

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2002, 12:38:07 : Search time 32.65 seconds  
(without alignments)  
1126.047 Million cell updates/sec

Title: US-09-884-948-1  
Perfect score: 1811  
Sequence: 1 DSDRVTPPAEPLDMPDPY.....ITFIPKSWNTAPDKVKQGP 331

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :	A_Geneseq_032802.*
1:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1980.DAT.*
2:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1981.DAT.*
3:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1982.DAT.*
4:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1983.DAT.*
5:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1984.DAT.*
6:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1985.DAT.*
7:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1986.DAT.*
8:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1987.DAT.*
9:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1988.DAT.*
10:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1989.DAT.*
11:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1990.DAT.*
12:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1991.DAT.*
13:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1992.DAT.*
14:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1993.DAT.*
15:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1994.DAT.*
16:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1995.DAT.*
17:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1996.DAT.*
18:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1997.DAT.*
19:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1998.DAT.*
20:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA1999.DAT.*
21:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA2000.DAT.*
22:	/SIDSL/gcgdata/hold-geneseq/geneqseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1811	100.0	331	20	AAV33662
2	1811	100.0	331	20	AAV33665
3	1811	100.0	331	20	AAW67770
4	1811	100.0	331	22	AAAB81161
5	1811	100.0	332	20	AAW67771
6	1811	100.0	332	21	AAAB12809
7	1811	100.0	345	15	AAAR49048
8	1811	100.0	406	13	AAAR22651
9	1811	100.0	407	22	AAAB97831
10	1811	100.0	407	22	AAAB81166
11	1811	100.0	408	15	AAAR49049

12	1802	99.5	331	13	AAAR22653	Transglutaminase (
13	1493	82.4	400	22	AAAB47007	Transglutaminase.
14	1487	82.1	330	22	AAAB81164	Transglutaminase r
15	1487	82.1	416	22	AAAB37830	S. cinnamomeum IFO
16	1465	80.9	331	17	AAAR51570	Transglutaminase
17	1391	76.8	328	20	AAV33666	Streptovorticilliu
18	162	8.9	39	17	AAAR98162	Transglutaminase N
19	140	7.7	30	17	AAAR98199	Peptide fragment o
20	130	7.2	26	17	AAAR98174	Peptide fragment o
21	124	6.8	25	17	AAAR98175	Peptide fragment o
22	119	6.6	23	17	AAAR98200	Peptide fragment o
23	116	6.4	23	17	AAAR98176	Peptide fragment o
24	114.5	6.3	889	22	AAAB55872	Drosophila melanog
25	112.5	6.2	1161	22	AAAB70419	Drosophila melanog
26	110	6.1	736	22	AAAB59845	AAV3A capsid prote
27	109	6.0	21	17	AAAR98202	Peptide fragment o
28	108	6.0	1005	21	AAAB43890	Human cancer assoc
29	108	6.0	1382	18	AAW31867	Human metastasis-a
30	107	5.9	736	22	AAAB59846	AAV3B capsid prote
31	106.5	5.9	736	21	AAV71167	Adeno-associated v
32	106.5	5.9	736	22	AAAB59847	AAV6 capsid protei
33	106	5.9	23	17	AAAR98188	Peptide fragment o
34	105	5.8	764	21	AAAG33024	Arabidopsis thalia
35	105	5.8	906	21	AAAG33023	Arabidopsis thalia
36	105	5.8	919	21	AAAG33022	Arabidopsis thalia
37	104	5.7	24	17	AAAR98169	Peptide fragment o
38	103	5.7	517	13	AAAR22904	1-Caldesmon. Gall
39	102	5.6	734	22	AAAB50326	Adeno-associated v
40	102	5.6	735	22	AAAG5792	Adeno-associated v
41	102	5.6	735	22	AAAS1508	Adeno-associated v
42	102	5.6	735	22	AAAB59844	AAV2 capsid protei
43	101	5.6	21	17	AAAR98163	Peptide fragment o
44	100	5.5	779	22	AAAG70899	C albicans apoptos
45	99.5	5.5	939	22	AAAB59143	Drosophila melanog

ALIGNMENTS

RESULT 1  
AAV33662  
ID AAV33662 standard; Protein; 331 AA.  
XX  
AC AAV33662;  
DT  
DT 07-JAN-2000 (first entry)  
XX  
XX Streptovorticillium S-8112 transglutaminase protein fragment.  
DE  
DE Transglutaminase; food industry; pharmaceutical industry; texture;  
KW Transglutaminase; food industry; pharmaceutical industry; texture;  
KW cosmetic industry; proteinaceous material; gel strength; viscosity;  
KW breaking strength; elasticity; taste.  
XX  
OS Streptovorticillium sp.  
XX  
XX DE19814860-A1.  
XX  
PD 07-OCT-1999.  
XX  
XX 02-APR-1998; 98DE-1014860.  
XX  
XX 02-APR-1998; 98DE-1014860.  
XX (FUCH/) FUCHSBAUER H.  
XX  
XX Fuchsbauser H, Pasternack R, Dorsch S, Otterbach J, Robenek I;  
PI Mainusch M, Dauscher C;  
XX  
XX WPI: 1999-552288/47.  
XX  
XX N-PSDB; AA223653.  
XX  
XX Bacterial transglutaminase polypeptides useful for polymerizing  
PT prepro-transglutam  
PT proteins, e.g. to modify the properties of food, pharmaceutical or

```

PT cosmetic products -
XX
XX Claim 1; Page 23-24; 44pp; German.
XX
CC This invention describes a novel bacterial transglutaminase polypeptide.
CC (I) can be used, e.g. in the food, pharmaceutical and cosmetic
CC industries, to polymerize proteinaceous materials in order to improve
CC their properties, e.g. texture, gel strength, breaking strength,
CC viscosity, elasticity or taste. (I) can also be used to immobilize
CC enzymes and antibodies. This sequence represents a transglutaminase
CC isolated from Streptovorticillum 8-8112.
XX
XX Sequence 331 AA:
SQ
Query Match 100.0%; Score 1811; DB 20; Length 331;
Best Local Similarity 100.0%; Pred. No. 2.3e-153;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DSDDRVTPPAEPLDAMPDPYPSPYSGRAETVVNNYIRKQQQVYSHRDGRKQOMTEQREW 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 dsddrvtpaepldrmpdypspyrsgraetvvnnyrkqqqvshrgkrqmqteeqrew 60
QY 61 SYGCVGVTVWNSGQYPTNRLAFASDFDEDFRNELKNRPSRGSTRAEFEGRVAKESFDEE 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 sygcvgvtnwsgqyptnrlafasdfdefrnelkngrpsrgsttraefegrvakesfdee 120
QY 121 KGFQRAREVASVMNRALENADHESAYLDNLKKELANGNDALRNEDARSFPYSALRNTPSF 180
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 kgfqrarevasvmnralenadenesayldnlkkelangndalrnedarspfysalrntpsf 180
QY 181 KERNGNHPDPRMKAVIVSKHFWSQDRSSSADRRKYGCDPAPPGTCLVDMSDRNT 240
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 kernngnhdpsrmkavivskhfwsgqdrssadrrkygdpdarppgtclvdmshrnt 240
QY 241 PRSPTSPGGGFVNFYGNFGAQTADABKTVMTHGNHYHAPNGSLGAMHYVESKFRNWE 300
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 prsptspgggfvnfygnfgaqtadabktvmtghnhyhapngslgamhyveskfrnwse 300
QY 301 GYSDFRGAYVITFTPKSNWTAPDKVKOGWP 331
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 301 qysdfdgayvitftfpkswnatapdkvkogwp 331

```

RESULT	2	
AAAY33665		
ID	AAAY33665	standard; Protein; 331 AA.
XX		
AC	AAAY33665;	
XX		
DT	07-JAN-2000	(first entry)
XX		
DE	Streptovorticillum mobaraense	transglutaminase protein fragment.
XX		
XX	Transglutaminase;	food industry; pharmaceutical industry; texture;
KW	cosmetic industry;	proteinaceous material; gel strength; viscosity;
KW	breaking strength;	elasticity; taste.
XX		
OS	Streptovorticillum mobaraense.	
XX		
PN	DE19814860-A1.	
XX		
PD	07-OCT-1999.	
XX		
PF	02-APR-1998;	98DE-1014860.
XX		
PR	02-APR-1998;	98DE-1014860.
XX		
PA	(FUCH/)	FUCHSBAUER H.
XX		
PI	Fuchsbaauer H,	Pasternack R, Dorsch S, Otterbach J, Robenek I;
PI	Mainusch M,	Dauscher C;
XX		

```

WIPI: 1999-552288/47.
XX Bacterial transglutaminase polypeptides useful for polymerizing
XX proteins, e.g. to modify the properties of food, pharmaceutical or
XX cosmetic products -
XX
XX Claim 12; Page 25-26; 44pp; German.
XX
XX This invention describes a novel bacterial transglutaminase polypeptide.
XX ({) can be used, e.g. in the food, pharmaceutical and cosmetic
XX industries, to polymerize proteinaceous materials in order to improve
XX their properties, e.g. texture, gel strength, breaking strength,
XX viscosity, elasticity or taste. ({) can also be used to immobilize
XX enzymes and antibodies. This sequence represents a transglutaminase
XX protein fragment isolated from Streptovorticillum mobaraense.
XX
XX Sequence 331 AA;
XX
Query Match 100.0%; Score 1811; DB 20; Length 331;
Best Local Similarity 100.0%; Pred. No. 2.3e-153;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0
QY 1 DSDRVRTPAEPLDRMPDPYRPSYGRAETVNNYIRKQQQVYSHRDGRKQOMTEQREWL 60
DB 1 dSDRVRTpPAePLdRmPdPyRpsYgrAetvnnYIrKqQqVYshRdGRkQOMteeqrEwL 60
QY 61 SYGCVGWTVNNSGYPTNRLAFASFDDEFRKVELKNGPRSGETRAEPEGRVAKESDEE 120
DB 61 syGcvGWtvNnSGyPtNrlAfASfdDEfRkVlKNGprSGetRAePEgrvAKesdEE 120
QY 121 KGFORAEVASVMNRALENAHDESAYLDMLKKELANGNDALRNEDARSPFYSA LRNTPSF 180
DB 121 kgfOrAEvASvmNRAlENaHdESAYldMLKKelANGndALrNEdARsPFySA LRntpsf 180
QY 181 KERNGNHDPFRMKAVIYKHFWSGGDRSSADKKRYGDPDAFRPAPOTGLVDKSRDNI 240
DB 181 kErngNhDpFRmkAViYkHfWsgGDRssADkKryGdPDaFRpAPoTGLvDKsRDnI 240
QY 241 PRSPTSPGEGFVNFDYGVWGAQTEADADKTVTHGHNHAPNGSLGAMHVYESKFRNWSE 300
DB 241 prsptspgEGfvnfdYgvwGaQteAdADkTVthGhnHApngSLgAMhvYEsKfrnwSe 300
QY 301 GYSDFDRGAVYITFIPKSWNTAPDKVQKGWP 331
DB 301 qYsdfdRGaVYITfIPkSWnTAPdKvQkgWp 331

```

RESULT	3	
AAW67770		
ID	AAW67770	standard; Protein; 331 AA.
XX		
AC	AAW67770;	
XX		
DT	01-APR-1999	(first entry)
XX		
DE	A transglutaminase	enzyme sequence.
XX		
KW	Transglutaminase;	microbial; gelled food; jelly; yogurt; cheese;
KW	cosmetic; meat quality;	microcapsule production; high thermal stability;
KW	carrier; immobilised	enzyme.
XX		
OS	Streptocorticillium	sp.
XX		
PN	EP889133-A2.	
XX		
PD	07-JAN-1999.	
XX		
PF	02-JUL-1998;	98EP-0112315.
XX		
PR	04-JUL-1997;	97JP-0180010.
XX		
PA	(AJIN )	AJINOMOTO CO INC.

```

XX MIWA T, Nakamura N, Seguro K, Yokoyama K;
PI WPI; 1999-062664/06.
XX N-PSDB; AAW81507.
XX New microbial transglutaminase with N-terminal aspartic acid deleted
PT - allowing high level recombinant production without added
PT methionine in E. coli, useful in production of gelled foods,
PT cosmetics etc.
XX Claim 1; Page 12-14; 56pp; English.
XX The present sequence represents a transglutaminase of Streptococcus
XX sp.. The specification describes a new microbial transglutaminase that
CC has the N-terminal aspartic acid of the present transglutaminase
CC deleted. Eliminating the N-terminal Asp from microbial transglutaminase
CC allows efficient removal of the terminal Methionine residue added when
CC the protein is expressed in Escherichia coli. The E. coli methionine
CC aminopeptidase acts well on Met-Ser but only poorly on Met-Asp, so
CC problems of antigenicity associated with Met-terminated proteins are
CC avoided. Recombinant transglutaminase is used to produce gelled foods
CC (Jellies, yogurt and cheeses) or cosmetics, to improve the quality of
CC meat, in the production of materials for microcapsules of high thermal
CC stability and as a carrier for immobilised enzymes.
XX Sequence 331 AA;
SQ
Query Match 100.0%; Score 1811; DB 20; Length 331;
Best Local Similarity 100.0%; Pred. No. 2.3e-153;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DSDDRVTTPAELDRMPDYPSPSYGRAETVYNNYIRKWOQVYSHRDGRKQMTTEQREWL 60
DB 1 dssdrvtppaeldrmpdpypspysgraetvnnvirkwqvyshrdgrkqmtteeqrewl 60
QY 61 SYGCVGTWVNSGOYPTNRLAFASDFDEKFNELKNGRPSGTRAEFGRVAKESFDEE 120
DB 1 dssdrvtppaeldrmpdpypspysgraetvnnvirkwqvyshrdgrkqmtteeqrewl 60
QY 121 KGFQAREVASVMNRALENAHDESAYLDNLKELANGNDALNEDARSFYSLRNTPSF 180
DB 121 kgfqrarevasvmnralenahdesayldnlkkelangndalnredarsphysalrntpsf 180
QY 181 KERNGNHDPSPRMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAGTGLVDMSRDRI 240
DB 181 kerngnhdpssrmkaviyskhfwsqdrssadkrkygdpdafrpapgtaglvdmsrdri 240
QY 241 PRSPTSPGEGFVNFYDYGWFGAQTADAKTWTGHNHYHAPNGSLGAMHYVESKFRNWE 300
DB 241 prsptspgegfvnfydygwfgaqteadaktwtgthgnhyhapngslgamhyveskfrnwse 300
QY 301 GYSDFDRGAYVITFIPKSNWNTAPDKVKQGW 331
DB 301 gysdfdrgayvitfipkswnntapdkvkqgw 331
RESULT 4
AAB81161
ID AAB81161 standard; protein: 331 AA.
XX AAB81161;
AC AAB81161;
XX 13-JUL-2001 (first entry)
XX Transglutaminase related protein SEQ ID 5.
DE Coryneform bacteria; transglutaminase; food processing.
XX Streptococcus thermophilus mobaraense.
OS Streptococcus thermophilus mobaraense.
XX WO200123591-A1.
PN

```

```

XX 05-APR-2001.
XX 29-SEP-2000; 2000WO-JP06780.
XX 30-SEP-1999; 99JP-0280098.
XX 28-JUN-2000; 2000JP-0194043.
XX (AJIN ) AJINOMOTO CO INC.
XX Kikuchi Y, Date M, Umezawa Y, Yokoyama K, Matsui H;
XX WPI; 2001-266172/27.
XX Efficient secretory production of foreign proteins e.g.
XX transglutaminase employing transformant coryneform bacterium, simply on
XX industrial scale with direct recovery for use in food processing and
XX pharmaceutical industry
XX Claim 22; Page 79-81; 151pp; Japanese.
XX This invention relates to a process for the production of a foreign
XX secretory protein through the construction of a recombinant coryneform
XX bacterium. The coryneform bacterium is transformed with an expression
XX construct in which DNA encoding a target foreign protein pro-structure is
XX ligated to the downstream region of DNA encoding the signal peptide
XX domain of a coryneform bacterial protein. Following transformation with
XX the vector, the bacterium is cultured, and the pro-peptide cleaved from
XX the expressed protein. Transglutaminases produced using this process are
XX useful in the food processing and pharmaceutical industries. The present
XX sequence represents a transglutaminase related protein, which can be
XX used in the method of the invention.
XX Sequence 331 AA;
SQ
Query Match 100.0%; Score 1811; DB 22; Length 331;
Best Local Similarity 100.0%; Pred. No. 2.3e-153;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 DSDDRVTTPAELDRMPDYPSPSYGRAETVYNNYIRKWOQVYSHRDGRKQMTTEQREWL 60
DB 1 dssdrvtppaeldrmpdpypspysgraetvnnvirkwqvyshrdgrkqmtteeqrewl 60
QY 61 SYGCVGTWVNSGOYPTNRLAFASDFDEKFNELKNGRPSGTRAEFGRVAKESFDEE 120
DB 61 sygcvgvtwvnsqyptnrlafasdfedrfknelkngprsgtraeefgrvakesfdee 120
QY 121 KGFQAREVASVMNRALENAHDESAYLDNLKELANGNDALNEDARSFYSLRNTPSF 180
DB 121 kgfqrarevasvmnralenahdesayldnlkkelangndalnredarsphysalrntpsf 180
QY 181 KERNGNHDPSPRMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAGTGLVDMSRDRI 240
DB 181 kerngnhdpssrmkaviyskhfwsqdrssadkrkygdpdafrpapgtaglvdmsrdri 240
QY 241 PRSPTSPGEGFVNFYDYGWFGAQTADAKTWTGHNHYHAPNGSLGAMHYVESKFRNWE 300
DB 241 prsptspgegfvnfydygwfgaqteadaktwtgthgnhyhapngslgamhyveskfrnwse 300
QY 301 GYSDFDRGAYVITFIPKSNWNTAPDKVKQGW 331
DB 301 gysdfdrgayvitfipkswnntapdkvkqgw 331
RESULT 5
AAB67771
ID AAB67771 standard; protein: 332 AA.
XX AAB67771;
AC AAB67771;
XX 01-APR-1999 (first entry)
XX

```

DE Protein encoded by high expression transglutaminase gene.  
 XX Transglutaminase; microbial; gelled food; jelly; yogurt; cheese;  
 KW cosmetic; meat quality; microcapsule production; high thermal stability;  
 KW carrier; immobilised enzyme.  
 XX  
 XX Synthetic.  
 OS Streptovorticillum sp.  
 PN EP889133-A2.  
 XX  
 XX 07-JAN-1999.  
 XX  
 XX 02-JUL-1998; 98EP-0112315.  
 XX  
 XX 04-JUL-1997; 97JP-0180010.  
 XX  
 XX (AJIN ) AJINOMOTO CO INC.  
 XX  
 XX Miwa T, Nakamura N, Seguro K, Yokoyama K;  
 PI WPI: 1999-062664/06.  
 DR N-PSDB; AAV81508.  
 DR  
 XX New microbial transglutaminase with N-terminal aspartic acid deleted  
 PT - allowing high level recombinant production without added  
 PT methionine in E. coli, useful in production of gelled foods,  
 PT cosmetics etc.  
 XX  
 XX Example 1; Page 18-23; 56pp; English.  
 PS  
 XX The present sequence is encoded by the high expression transglutaminase  
 XX gene present in plasmid pTAMG-02. The gene is derived from  
 CC Streptovorticillum sp., and is codon altered, using oligonucleotides  
 CC AAV81521-60, for expression in Escherichia coli. The specification  
 CC describes a new microbial transglutaminase that has the N-terminal  
 CC aspartic acid of transglutaminase deleted. Eliminating the N-terminal  
 CC Asp from microbial transglutaminase allows efficient removal of the  
 CC terminal Met residue added when the protein is expressed in E. coli.  
 CC The E. coli methionine aminopeptidase acts well on Met-Ser but only  
 CC poorly on Met-Asp, so problems of antigenicity associated with  
 CC Met-terminated proteins are avoided. Recombinant transglutaminase is  
 CC used to produce gelled foods (jellies, yogurt and cheeses) or cosmetics,  
 CC to improve the quality of meat, in the production of materials for  
 CC microcapsules of high thermal stability and as a carrier for immobilised  
 CC enzymes.  
 XX  
 XX Sequence 332 AA;  
 SQ  
 Query Match 100.0%; Score 1811; DB 20; Length 332;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-153;  
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DSDRVTTPAEPLDRMPDPYRPSYGRAETVNNYIRKQOQVYSHRDGRKQOQTEQREWL 60  
 DB 2 dsddrvtpaepldrmpdpypsygraetvnnvyrkwwqvysyhrdgrkqmtteqrewl 61  
 QY 61 SYGCVGVTVWNSGQYPTNRLAFASDFEDRFKNELKNRPSRGETRAEFGRVAKESFDEE 120  
 DB 62 sygcvgvttwnsgqyptnrlafasfdeedrfknelkngrpsgetraefgrvakesfdee 121  
 QY 121 KGFQREAVASVMNRALENAHDESAYLDNLKKELANGNDALRNEDARSPFYSA LRNTPSF 180  
 DB 122 kgfqrarevasvmnralenahdesayldnlkkelangndalrnedarspfysalrntpsf 181  
 QY 181 KERNGNHDSRMAVYIKHFSGGDRSSADKRKYGDPDAFRPAGTGLVDMRSRDN 240  
 DB 182 kerngnhdsrmkavlyskhfsggdrssadkrkygdpdfrpaggltglvdmrsdrni 241  
 QY 241 PRSTSPGEGFVNDYGFCAQTEADAKTWTWTHGHYHAPNGLGAMHVYESKPRNWE 300  
 DB 242 preptspegfvndygfcaqteadaktwtwhghnyhapngslgamhvyeskfwnwe 301

QY 301 GYSDFDRGAVYVIFIPKSWNTAPDKVKQGW 331  
 DB 302 gysdfdrgayvifipkswntapdkvkqgwp 332  
 ID AAB12809 standard; Protein; 332 AA.  
 XX AAB12809;  
 AC AAB12809;  
 XX 24-NOV-2000 (first entry)  
 DT Transglutaminase protein sequence SEQ ID NO:1.  
 DE Transglutaminase; gelled food; jelly; yogurt; gelled cosmetic;  
 KW cheese.  
 KW Unidentified.  
 OS WO200040706-A1.  
 PN 13-JUL-2000.  
 PD 24-DEC-1999; 99WO-JP07250.  
 PF 28-DEC-1998; 98JP-0373131.  
 XX (AJIN ) AJINOMOTO CO INC.  
 XX Yokoyama K, Ono K, Ejima D;  
 PI WPI: 2000-475826/41.  
 DR N-PSDB; AAA73025.  
 DR  
 XX Production of active transglutaminase from denatured enzyme by  
 PT two-stage refolding process for industrial production of active enzyme  
 PT for use in food production -  
 XX Disclosure; Page 48-50; 74pp; Japanese.  
 PS  
 XX The present invention describes a method for producing active  
 CC transglutaminase from denatured enzyme. The method comprises: (i) forming  
 CC an intermediate structure of the enzyme having transglutaminase activity  
 CC under acidic conditions in an aqueous medium; and (ii) forming a higher-  
 CC level structure of the enzyme having transglutaminase activity under  
 CC neutral conditions in an aqueous medium. The method can be used for  
 CC industrial production of active transglutaminase from denatured material  
 CC (such as recombinant transglutaminase) which can be used in the food  
 CC industry for the production of gelled foods such as jellies, yoghurts  
 CC and cheeses, and for the production of gelled cosmetics. The present  
 CC sequence represents a transglutaminase which is used in the  
 CC exemplification from the present invention.  
 XX  
 XX Sequence 332 AA;  
 SQ  
 Query Match 100.0%; Score 1811; DB 21; Length 332;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-153;  
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DSDRVTTPAEPLDRMPDPYRPSYGRAETVNNYIRKQOQVYSHRDGRKQOQTEQREWL 60  
 DB 2 dsddrvtpaepldrmpdpypsygraetvnnvyrkwwqvysyhrdgrkqmtteqrewl 61  
 QY 61 SYGCVGVTVWNSGQYPTNRLAFASDFEDRFKNELKNRPSRGETRAEFGRVAKESFDEE 120  
 DB 62 sygcvgvttwnsgqyptnrlafasfdeedrfknelkngrpsgetraefgrvakesfdee 121  
 QY 121 KGFQREAVASVMNRALENAHDESAYLDNLKKELANGNDALRNEDARSPFYSA LRNTPSF 180  
 DB 122 kgfqrarevasvmnralenahdesayldnlkkelangndalrnedarspfysalrntpsf 181

```

QY 181 KERNGNHDPSPRMKAVIYKHFWSGQDRSSADKRRKYGDPDAFRPAPGTGLVDMSRDRI 240
    |||||||
Db 182 kerngnhdpSRmkaviskhfwsgqdrssadkrrkygdpdfrpapgtglvdmSRdrni 241
    |||||||
QY 241 PRSPTSPGEGFVNFYDYGWFGAQTADAKTVTHGNHYHAPNGSLGAMHYESKFRNWSE 300
    |||||||
Db 242 prsptspgEgfvnfydgwfgaqtEadaktvthgnhyhapngslgamhyesKfrnwse 301
    |||||||
QY 301 GYSDFDRGAYVITFIPKSWNTAPDKVKQGW 331
    |||||||
Db 302 gysdfdrGAYvitfipkswntapdkvkqgw 332
    |||||||

RESULT 7
ID AAR49048 standard; Protein; 346 AA.
XX AAR49048;
XX 20-SEP-1994 (first entry)
XX Bacterial transglutaminase.
XX Bacterial transglutaminase; BTG; expression;
KW active; inactive; inclusion body.
XX JP06030771-A.
XX 08-FEB-1994.
XX 14-JUL-1992; 92JP-0187038.
XX 14-JUL-1992; 92JP-0187038.
XX (AJIN ) AJINOMOTO KK.
XX WPI; 1994-079294/10.
XX N-PSDB; AAQ55983.
XX Prepn. of bacterial trans-glutaminase in large amts. - by
PT expression of fusion protein in E. coli bacterial
PT trans-glutaminase
XX Claim 1; Page 7-8; 13pp; Japanese.
XX A fused protein contains amino acids 16-346 of BTG (AAQ55983)
CC and a hydrophilic peptide at the amino terminal.
CC Expression of DNA encoding this protein in E. coli allow large
CC scale prodn. of BTG. An active BTG can be prepd. from the
CC inactive fused protein inclusion body.
XX Sequence 346 AA;

Query Match 100.0%; Score 1811; DB 15; Length 346;
Best Local Similarity 100.0%; Pred. No. 2.4e-153;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTTPAEPLDRMDPPRPSYGRAETVNNYKRWQVYSHRGRKQOMTEQEWL 60
    |||||||
Db 16 dsddrvtpPaepldrmpdprrpsYgraetvnnYirkwqvYshrdgrkqmqteedrewl 75
    |||||||
QY 61 SYGCVGVTVNWSGQYPTNRLAFASFDREPKELKNGRPSRGETRAEFEGRVAKESFDDE 120
    |||||||
Db 76 sygcvgvTVnwsGqyptnrlafasfdredrFknelkngprsrgetraefegrvakesfdee 135
    |||||||
QY 121 KGFORAREVASVNNRLENLHDSAYLDNLKKELANGDNLARNEDARSPFYSALRNTPSF 180
    |||||||
Db 136 kgfgrarevasvnnrAlenahdesayldnlkkelangdnlarNedarSPfysalrntpsf 195
    |||||||
QY 181 KERNGNHDPSPRMKAVIYKHFWSGQDRSSADKRRKYGDPDAFRPAPGTGLVDMSRDRI 240
    |||||||

```

```

Db 196 kerngnhdpSRmkaviskhfwsgqdrssadkrrkygdpdfrpapgtglvdmSRdrni 255
    |||||||
QY 241 PRSPTSPGEGFVNFYDYGWFGAQTADAKTVTHGNHYHAPNGSLGAMHYESKFRNWSE 300
    |||||||
Db 256 prsptspgEgfvnfydgwfgaqtEadaktvthgnhyhapngslgamhyesKfrnwse 315
    |||||||
QY 301 GYSDFDRGAYVITFIPKSWNTAPDKVKQGW 331
    |||||||
Db 316 gysdfdrGAYvitfipkswntapdkvkqgw 346
    |||||||

RESULT 8
ID AAR22651 standard; Protein; 406 AA.
XX AAR22651;
XX 09-OCT-1992 (first entry)
XX Transglutaminase.
XX BTG; acyl rearrangement; deamination.
XX Actinomycetes and Streptovorticillium.
XX Key Location/Qualifiers
FT Peptide 1..75
FT Protein /label= sig_peptide
FT /label= mat_BTG
XX EP481504-A.
XX 22-APR-1992.
XX 18-OCT-1991; 91EP-0117813.
XX 19-OCT-1990; 90JP-0282566.
XX (AJIN ) AJINOMOTO CO INC.
XX (AMAN ) AMANO PHARM KK.
XX Ando K, Arafuka S, Koikeda S, Matsui H, Takagi H, Washizu K;
XX WPI; 1992-133808/17.
XX N-PSDB; AAQ24197 and AAQ24201.
XX DNA fragment encoding trans:glutaminase - is inserted into
PT vector, e.g. pMJ053-BTG, for protein expression
XX Disclosure; Page 42 and 44; 55 pp; English.
XX The mature transglutaminase enzyme (BTG) can be derived from two
CC different species, Streptovorticillium and Actinomycetes. It
CC catalyzes an acyl rearrangement reaction of a gamma-carboxamide gp.
CC of glutamine. It introduces intra- or intermolecular formation of
CC epsilon-(gamma-Gln)-Lys cross-linking when an epsilon-amino gp. of
CC a Lys residue acts as an acyl receptor. When water acts as an acyl
CC acceptor the enzyme accelerates the conversion of Gln residues to Glu
CC residues by deamination.
CC The enzyme is used in the prodn. of gelled foods, gelled cosmetics,
CC yogurt, gelatin, cheese etc. It is also used in the prodn. of
CC thermally stable materials such as microcapsules and carriers of
CC immobilized enzymes. The DNA sequence given allows the prodn. of
CC BTG efficiently and in large quantity.
XX Sequence 406 AA;

Query Match 100.0%; Score 1811; DB 13; Length 406;
Best Local Similarity 100.0%; Pred. No. 3e-153;
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 DSDRVTTPAEPLDRMPDPYPSYGRAETVNNYIRKQOQVYSHRDRGKQOQTEBQREWL 60  
 Db 76 dsddrvtpaepldrmpdpypsygraetvnnnyirkwqvyshrdgrkgmtceeqrewl 135  
 QY 61 SYGCVGVTWVNSGOYPTNRLAFASFEDEDFRKNELKNGRPSGETRAEFEGRVAKESFDEE 120  
 Db 136 sygcvgvtwvnsqyptnrlafasfededfrknelkngprsgsetraefegrvakesfdee 195  
 QY 121 KGFORAREVASVMNRALENHADESAYLDNKKELANGNDALRNEDARSPPYSALRNTPSF 180  
 Db 196 kgfgrarevasvmnralenahdesayldnlkkelangndalrnedarspfysalrntpsf 255  
 QY 181 KERNGNHDPSSRMKAVIYSKHFWSGQDRSSSADKRYGDPDAFRPAGTGLVDMXSRDRI 240  
 Db 256 kerngnhdpssrmkaviyskhfwsqgdrssssadkrkygdpdafpapggtglvdmxsrdrni 315  
 QY 241 PRSPSPGEGFVNFYDGNFGAQTADADKTVTHGNHYHAPNGSLGAMHYVESKFRNWE 300  
 Db 316 prspspggefvnfydgnfgaqtadadktvthgnhyhapngslgamhyveskfrnwse 375  
 QY 301 GYSDFDRGAYVITFIPKSWNTAPDKVKQGW 331  
 Db 376 gysdfdrgayvitfipkswntapdkvkqgw 406

## RESULT 9

AAB97831  
 ID AAB97831 standard; Protein; 407 AA.

XX AAB97831;

DT 09-AUG-2001 (first entry)

DE S. mobaraense IFO 13819 transglutaminase protein SEQ ID NO:4.

XX Streptovorticillium cinnamonum IFO 12852; Streptomyces; actinomycete;

KW Streptovorticillium mobaraense IFO 13819; transglutaminase.

XX Streptovorticillium mobaraensis.

OS WC200129187-A1.

PN 26-APR-2001.

PD 13-OCT-2000; 2000WO-JP07135.

PF 18-OCT-1999; 99JP-0295649.

PR (AJIN ) AJINOMOTO CO INC.

XX Taguchi S, Momose H;

PI WPI; 2001-300330/31.

XX N-PSDB; AAH20188.

XX Streptomyces sp. carrying an actinomycete-derived gene and promoter for producing high yields of transglutaminase

PS Disclosure; Page 33-36; 41pp; Japanese.

XX The present invention describes a Streptomyces sp. containing a gene construct comprising actinomycete-derived transglutaminase gene and promoter. Also described are methods for producing pro-transglutaminase and active transglutaminase. The gene construct can be used in the production of large amounts of transglutaminase. The present sequence represents Streptovorticillium mobaraense IFO 13819 transglutaminase, which is given in the exemplification of the present invention.

XX Sequence 407 AA;

Query Match 100.0%; Score 1811; DB 22; Length 407;  
 Best Local Similarity 100.0%; Pred. No. 3e-153;

Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DSDRVTTPAEPLDRMPDPYPSYGRAETVNNYIRKQOQVYSHRDRGKQOQTEBQREWL 60  
 Db 76 dsddrvtpaepldrmpdpypsygraetvnnnyirkwqvyshrdgrkgmtceeqrewl 136  
 QY 61 SYGCVGVTWVNSGOYPTNRLAFASFEDEDFRKNELKNGRPSGETRAEFEGRVAKESFDEE 120  
 Db 137 sygcvgvtwvnsqyptnrlafasfededfrknelkngprsgsetraefegrvakesfdee 196  
 QY 121 KGFORAREVASVMNRALENHADESAYLDNKKELANGNDALRNEDARSPPYSALRNTPSF 180  
 Db 197 kgfgrarevasvmnralenahdesayldnlkkelangndalrnedarspfysalrntpsf 256  
 QY 181 KERNGNHDPSSRMKAVIYSKHFWSGQDRSSSADKRYGDPDAFRPAGTGLVDMXSRDRI 240  
 Db 257 kerngnhdpssrmkaviyskhfwsqgdrssssadkrkygdpdafpapggtglvdmxsrdrni 316  
 QY 241 PRSPSPGEGFVNFYDGNFGAQTADADKTVTHGNHYHAPNGSLGAMHYVESKFRNWE 300  
 Db 317 prspspggefvnfydgnfgaqtadadktvthgnhyhapngslgamhyveskfrnwse 376  
 QY 301 GYSDFDRGAYVITFIPKSWNTAPDKVKQGW 331  
 Db 377 gysdfdrgayvitfipkswntapdkvkqgw 407

## RESULT 10

AAB81166

ID AAB81166 standard; Protein; 407 AA.

XX AAB81166;

DT 13-JUL-2001 (first entry)

DE Prepro-transglutaminase amino acid sequence.

KW Coryneform bacteria; transglutaminase; food processing.

XX Streptovorticillium mobaraense.

OS WC200123591-A1.

PN 05-APR-2001.

PD 29-SEP-2000; 2000WO-JP06780.

PF 30-SEP-1999; 99JP-0280098.

PR 28-JUN-2000; 2000JP-0194043.

XX (AJIN ) AJINOMOTO CO INC.

PI Kikuchi Y, Date M, Umezawa Y, Yokoyama K, Matsui H;

XX WPI; 2001-266172/27.

XX N-PSDB; AAF86283.

XX Efficient secretory production of foreign proteins e.g. transglutaminase employing transformant coryneform bacterium, simply on industrial scale with direct recovery for use in food processing and pharmaceutical industry

PS Example 1; Page 90-93; 151pp; Japanese.

XX This invention relates to a process for the production of a foreign secretory protein through the construction of a recombinant coryneform bacterium. The coryneform bacterium is transformed with an expression construct in which DNA encoding a target foreign protein pro-structure is ligated to the downstream region of DNA encoding the signal peptide domain of a coryneform bacterial protein. Following transformation with the vector, the bacterium is cultured, and the pro-peptide cleaved from the expressed protein. Transglutaminases produced using this process are useful in the food processing and pharmaceutical industries. The present

CC sequence prepro-transglutaminase. The protein is used in an example  
 CC illustrating the method of the invention.

SQ Sequence 407 AA;  
 Query Match 100.0%; Score 1811; DB 22; Length 407;  
 Best Local Similarity 100.0%; Pred. No. 3e-153;  
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DSDRVTPPAEPLDRMPDPYRPSYGRAETVYNNYIRKQOQYSHRDGRKQOQTEQREW 60  
 DB 77 dsdrrvtpaepldrmpdpypysygraetvnnnyirkwqyshrdgrkqgnteeqrewl 136  
 QY 61 SYGCVGTWVNSQYPTNRLAFASDFEDFRKNEKNGRPSRGETRAEFGRVAKESFDEE 120  
 DB 137 sygcvgvtwnsgyptnrlafasdfedfrknekngrpsrgetraefgrvakesfdee 196  
 QY 121 KGFQAREVASVMNRALENAHDESAYLDNLKKELANGNDALRNEDARSFYSALRNTPSF 180  
 DB 197 kgfqrarevasvmnrallenahdesayldnlkkelangndalrnedarspfysalrntpsf 256  
 QY 181 KERNGNHDPSRMKAVIYKSHFWSGQDRSSADKRYGDPDAFRPAPGTGLVDMSRDRNI 240  
 DB 257 kerngnhdpsrmkaviyskhfwsqdrssadkrygdpdafpapggtglvdmrdrni 316  
 QY 241 PRSPTSPGSEGVNFYDYGWFGAQTEADADKTWTHGNHYHAPNGSLGAMHYVESKFRNWSE 300  
 DB 317 prsptspgegvnfdygwfgaqteadadktwthgnhyhapngslgamhyveskfrnwse 376  
 QY 301 GYSDFDRGAYVITFIPKSWNTAPDKYKQGW 331  
 DB 377 gysdfdrgayvitfipkswntapdkvkqgwp 407

RESULT 11  
 AAR49049  
 ID AAR49049 standard; Protein; 408 AA.  
 AC AAR49049;  
 XX  
 XX  
 DT 20-SEP-1994 (first entry)  
 XX  
 XX BTG-contg. sequence.  
 XX  
 XX Bacterial transglutaminase; BTG; expression;  
 KW active; inactive; inclusion body.  
 XX  
 XX JF06030771-A.  
 XX  
 XX 08-FEB-1994.  
 XX  
 XX 14-JUL-1992; 92JP-0187038.  
 XX  
 XX 14-JUL-1992; 92JP-0187038.  
 XX  
 XX (AJIN ) AJINOMOTO KK.  
 XX  
 XX WPI; 1994-079294/10.  
 DR N-PSDB; AA055984.  
 XX  
 XX Prepn. of bacterial trans-glutaminase in large amts. - by  
 PT expression of fusion protein in E. coli bacterial  
 PT trans-glutaminase  
 XX  
 XX Disclosure; Page 8-10; 13pp; Japanese.  
 XX  
 XX A fused protein contains amino acids 16-346 of BTG (AAQ55983)  
 CC and a hydrophilic peptide at the amino terminal.  
 CC Expression of DNA encoding this protein in E. coli allow large  
 CC scale prodn. of BTG. An active BTG can be prepd. from the  
 CC inactive fused protein inclusion body.  
 CC  
 XX

SQ Sequence 408 AA;  
 Query Match 100.0%; Score 1811; DB 15; Length 408;  
 Best Local Similarity 100.0%; Pred. No. 3e-153;  
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 DSDRVTPPAEPLDRMPDPYRPSYGRAETVYNNYIRKQOQYSHRDGRKQOQTEQREW 60  
 DB 78 dsdrrvtpaepldrmpdpypysygraetvnnnyirkwqyshrdgrkqgnteeqrewl 137  
 QY 61 SYGCVGTWVNSQYPTNRLAFASDFEDFRKNEKNGRPSRGETRAEFGRVAKESFDEE 120  
 DB 138 sygcvgvtwnsgyptnrlafasdfedfrknekngrpsrgetraefgrvakesfdee 197  
 QY 121 KGFQAREVASVMNRALENAHDESAYLDNLKKELANGNDALRNEDARSFYSALRNTPSF 180  
 DB 198 kgfqrarevasvmnrallenahdesayldnlkkelangndalrnedarspfysalrntpsf 257  
 QY 181 KERNGNHDPSRMKAVIYKSHFWSGQDRSSADKRYGDPDAFRPAPGTGLVDMSRDRNI 240  
 DB 258 kerngnhdpsrmkaviyskhfwsqdrssadkrygdpdafpapggtglvdmrdrni 317  
 QY 241 PRSPTSPGSEGVNFYDYGWFGAQTEADADKTWTHGNHYHAPNGSLGAMHYVESKFRNWSE 300  
 DB 318 prsptspgegvnfdygwfgaqteadadktwthgnhyhapngslgamhyveskfrnwse 377  
 QY 301 GYSDFDRGAYVITFIPKSWNTAPDKYKQGW 331  
 DB 378 gysdfdrgayvitfipkswntapdkvkqgwp 408

RESULT 12  
 AAR22653  
 ID AAR22653 standard; Protein; 331 AA.  
 XX  
 AC AAR22653;  
 XX  
 DT 09-OCT-1992 (first entry)  
 XX  
 XX Transglutaminase (expressed in E. coli).  
 DE  
 XX BTG; acyl rearrangement; deamination; yeast; actinomycetes.  
 KW  
 XX Synthetic.  
 OS  
 XX EP481504-A.  
 XX  
 XX 22-APR-1992.  
 XX  
 XX 18-OCT-1991; 91EP-0117813.  
 XX  
 XX 19-OCT-1990; 90JP-0282566.  
 XX  
 XX (AJIN ) AJINOMOTO CO INC.  
 PA (AMANO ) AMANO PHARM KK.  
 XX  
 XX Ando K, Arafuka S, Koikeda S, Matsui H, Takagi H, Washizu K;  
 XX  
 XX WPI; 1992-133808/17.  
 DR N-PSDB; AAQ24207.  
 XX  
 XX DNA fragment encoding trans:glutaminase - is inserted into  
 PT vector, e.g. PnJ1053-BTG, for protein expression  
 PT  
 XX Disclosure; Page 3; 55pp; English.  
 PS  
 XX The protein sequence given has transglutaminase (BTG) activity. The  
 CC DNA encoding this protein has a base sequence which can be used  
 CC suitably in an expression system using E. coli or yeast as a host.  
 CC The base sequence can be compared to those given in AAQ24197 and  
 CC AAQ24200 which are derived from Actinomycetes sp.  
 CC BTG catalyses an acyl rearrangement reaction of a gamma-carboxamide  
 CC

CC gp. of glutamine. It introduces intra- or intermolecular formation  
 CC of epsilon-(gamma-Gln)-Lys cross-linking when an epsilon-amino gp.  
 CC of a Lys residue acts as an acyl receptor. When water acts as an  
 CC acyl acceptor the enzyme accelerates the conversion of Gln residues  
 CC to Gln residues by deamination.  
 CC The enzyme is used in the prodn. of gelled foods, gelled cosmetics,  
 CC yogurt, gelatin, cheese etc. It is also used in the prodn. of  
 CC thermally stable materials such as microcapsules and carriers of  
 CC immobilized enzymes. The DNA sequence given allows the prodn. of  
 CC BTG efficiently and in large quantity.  
 XX Sequence 331 AA;

SQ  
 Query Match 99.5%; Score 1802; DB 13; Length 331;  
 Best Local Similarity 99.7%; Pred. No. 1.4e-152;  
 Matches 330; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 DSDDRVTPAEPLDRMPDPYRPSYGRAETVNNYIRKQOQVYSHRDKRQKQWTEQREWL 60  
 Dd 1 dddrvtpaepldrmpdpypysygraetvnnyrkqvyshrdgrkqgqteegrewl 60  
 Qy 61 SYGCVGVTVWNSGQYPTNRLAFASDEDRFKNELKNGRPSRGETRAEFGRVAKESFDEE 120  
 Dd 61 sygcvgv tvwnsgqyptnrlafasdedr fknkngprsrgetraefgrvakesfdee 120  
 Qy 121 KGFQRAREVASVNNRALENADHESAYLDNLKELANGNDALRNEDARSPFYSALRNTPSF 180  
 Dd 121 kgfqrarevasvnnralenahdesayldnlkkelangndalrnedarspfysalrntpsf 180  
 Qy 181 KERNGGNHDP SRMKAVIYSKHFWSGQDRSSADKRYKGDPPAFRPAQGTGLVDMSRDRNI 240  
 Dd 181 kerngnghdp srmkaviyskhfswgdrssadkrykqgdpafrpapggtglvdmrdrni 240  
 Qy 241 PRSPTSPGSGGFVNDYGWFGAGTEADADKTVTHGNHGHAPNGSLGAMHYVESKFRNWE 300  
 Dd 241 prsptspgsggf vndygwfgagteadadk tvthgnhghapngslgamhyveskfrnwse 300  
 Qy 301 GYSDFORGAYVITFIPKSWNTAPDKVKQGP 331  
 Dd 301 gysdforgayvitfipkswntapdkvkqgp 331

RESULT 13  
 AAB47007  
 ID AAB47007 standard; Protein; 400 AA.

AC AAB47007;

DT 22-MAR-2001 (first entry)

DE Transglutaminase.

XX Transglutaminase; acyl transfer; glutamine; ATCC 27446; food;  
 KW cosmetic; pharmaceutical; gel; artificial skin.

OS Streptovorticillium mobaraense.

XX WO200070026-A1.

PN 23-NOV-2000.

XX 10-MAY-2000; 2000WO-US12601.

XX 14-MAY-1999; 99US-0134158.

XX (WISC ) WISCONSIN ALUMNI RES FOUND.

XX Damodaran S;

PI WPI: 2001-070774/08.

DR N-PSDB; AAC85251.

XX

PT New transglutaminase enzyme from Streptovorticillium mobaraense for  
 PT e.g. use in gelled products and the production of artificial skin  
 XX Claim 8; Page 44-45; 49pp; English.

XX This sequence represents a transglutaminase which catalyses the acyl  
 CC transfer of the gamma-carboxamide group of a glutamine residue in  
 CC a peptide or protein chain independently of calcium ions and which  
 CC has an activity at pH 9.0 that is 40 % or greater than its activity  
 CC at pH 7.0. The transglutaminase was isolated from Streptovorticillium  
 CC mobaraense strain ATCC (American Type Culture Collection) No. 27446.  
 CC The transglutaminases catalyse an acyl transfer reaction of a  
 CC gamma-carboxamide group of a glutamine residue and a primary amine of  
 CC a peptide. When the epsilon-amino group of a lysine residue functions as  
 CC the acyl acceptor, intramolecular and intermolecular cross-linking  
 CC occurs. When water functions as the acyl acceptor, transglutaminase  
 CC converts glutamine residues in glutamic acid residues by deamidation.  
 CC The cross-linking reaction is useful in the food, cosmetic and  
 CC pharmaceutical industries. Transglutaminase can gel protein, making it  
 CC useful in production of gelled food, gelled cosmetics, gelatins,  
 CC yogurt, cheese and other products. The enzyme can also be used to make  
 CC thermally stable materials such as microcapsules or carriers of  
 CC immobilized enzymes. The cross-linking reaction is also potentially  
 CC useful in production of artificial skin. The transglutaminase coding  
 CC sequence may be used as probes for detecting the presence and/or  
 CC expression of streptovorticillium transglutaminase genes, or for  
 CC identifying related genes from other microbial species. They may  
 CC also be used to produce large quantities of the enzyme.  
 CC The new transglutaminase has a higher activity, both at pH  
 CC 7.0 and pH 9.0, when compared to commercially available  
 CC transglutaminases. It has a different susceptibility to several  
 CC commonly used inhibitors, when compared to previously isolated  
 CC transglutaminases. These properties broaden the range of application  
 CC in which the new enzyme can be used.

XX Sequence 400 AA;

Query Match 82.4%; Score 1493; DB 22; Length 400;  
 Best Local Similarity 82.3%; Pred. No. 7.3e-125;  
 Matches 269; Conservative 25; Mismatches 33; Indels 0; Gaps 0;

Qy 5 RVTPPAEPLDRMPDPYRPSYGRAETVNNYIRKQOQVYSHRDKRQKQWTEQREWLISYGC 64

Dd 74 retppaepldrmpdayrarggrattvnnyrkqvyshldgkqgqteegreklsygc 133

Qy 65 VGVTVWNSGQYPTNRLAFASDEDRFKNELKNGRPSRGETRAEFGRVAKESFDEEKGQ 124

Dd 134 vgv tvwnsgqyptnrlafsfddedkykndlkntrplagetraefgrakafdegkfk 193

Qy 125 RAREVASVNNRALENADHESAYLDNLKELANGNDALRNEDARSPFYSALRNTPSFKERN 184

Dd 194 rardvaslmmkalenahdegayldkltelnnndallhedrsnfysalrntpsfkrd 253

Qy 185 GGNHDP SRMKAVIYSKHFWSGQDRSSADKRYKGDPPAFRPAQGTGLVDMSRDRNIPRSP 244

Dd 254 ggn ydpsr mkaviyskhfswgqdrsgsdkrkygdpaeafprdggtglvdmkskdnprsp 313

Qy 245 TSPGEGGFVNDYGWFGAGTEADADKTVTHGNHGHAPNGSLGAMHYVESKFRNWESEYSD 304

Dd 314 anpgewwnfygwfgagteadadk tlvthgdnhyhapngdlgpmhvyesfrnwseayad 373

Qy 305 FDRGAYVITFIPKSWNTAPDKVKQGP 331

Dd 374 fdrgtymialipkswntapakvkqgp 400

RESULT 14

AAB81164

ID AAB81164 standard; protein; 330 AA.

XX AC AAB81164;

XX

DT 13-JUL-2001 (first entry)  
 XX Transglutaminase related protein SEQ ID 43.  
 DE Coryneform bacteria; transglutaminase; food processing.  
 XX Streptovorticillum cinnamomeum.  
 XX WO200123591-A1.  
 PN 05-APR-2001.  
 XX 29-SEP-2000; 2000WO-JP06780.  
 XX 30-SEP-1999; 99JP-0280098.  
 PR 28-JUN-2000; 2000JP-0194043.  
 XX (AJIN ) AJINOMOTO CO INC.  
 PA Kikuchi Y, Date M, Umezawa Y, Yokoyama K, Matsui H;  
 PI WPI; 2001-255172/27.  
 DR  
 XX Efficient secretory production of foreign proteins e.g.  
 PT transglutaminase employing transformant coryneform bacterium, simply on  
 PT industrial scale with direct recovery for use in food processing and  
 PT pharmaceutical industry  
 XX Claim 24; Page 131-133; 151pp; Japanese.  
 PS  
 CC This invention relates to a process for the production of a foreign  
 CC secretory protein through the construction of a recombinant coryneform  
 CC bacterium. The coryneform bacterium is transformed with an expression  
 CC construct in which DNA encoding a target foreign protein pro-structure is  
 CC ligated to the downstream region of DNA encoding the signal peptide  
 CC domain of a coryneform bacterial protein. Following transformation with  
 CC the vector, the bacterium is cultured and the pro-peptide cleaved from  
 CC the expressed protein. Transglutaminases produced using this process are  
 CC useful in the food processing and pharmaceutical industries. The present  
 CC sequence represents a transglutaminase related protein, which can be  
 CC used in the method of the invention.  
 XX Sequence 330 AA;  
 SQ  
 Query Match 82.1%; Score 1487; DB 22; Length 330;  
 Best Local Similarity 81.5%; Pred. NO. 1.9e-124;  
 Matches 269; Conservative 26; Mismatches 35; Indels 0; Gaps 0;  
 QY 2 SDDRVTTPAEPDLPDPYRPSYGRAETVNNYIRKWOQVYSHRDGRKQOMTEQREWLS 61  
 DB 1 sddretppaepldrmpayrayaggrattvnnnyirkwqvyshrdgkqgkqteeqrekl 60  
 QY 62 YGCVGVTVWNSGOYPTNRLAFASFDEDFRKNELKNGRPRSGETRAEFEGRVAKESFDEEK 121  
 DB 61 ygcvgvtywvnsqgpyptnrlafasfdenkyndkntsprdpdtraefegriakgsfdegk 120  
 QY 122 GFORAREVASVMNRALENAHDEGSAYLDNLKELANGNDALRNEDARSFPYSALRNTSPSK 181  
 DB 121 gfkardvasvmnkalenahdegtyinnlktelnnndallredsrnsfysalrntpsfk 180  
 QY 182 ERNGGNHDPGRMKAVIYSKHFWSGQDRSSADRKRYGDPDAFRPAGTGLVDMSSDRNP 241  
 DB 181 erdgngnydpkrmkavlyskhfwsgdrgsgskrykgydpeafpddggtglvdmksdrsp 240  
 QY 242 RSPTSPGCEGVNFDYGFAGQATEADAKTVTHGNHYPAPNGSLGAMHVESKFRNWSG 301  
 DB 241 rspakpgegvnfdygfagqateadaktwthghyapnsglghpmhveskfrkwsag 300  
 QY 302 YSDFDRGAYVITFIPKSWNTAPDKVKQGW 331  
 DB 301 yadfdrgayvitfipkswntapakvegwp 330

RESULT 15  
 AAB97830  
 ID AAB97830 standard; Protein; 416 AA.  
 XX  
 AC AAB97830;  
 XX 09-AUG-2001 (first entry)  
 DT  
 DE S. cinnamomeum IFO 12852 transglutaminase protein SEQ ID NO:2.  
 XX Streptovorticillum cinnamomeum IFO 12852; Streptomyces; actinomycete;  
 KW Streptovorticillum mobarsense IFO 13819; transglutaminase.  
 XX Streptovorticillum cinnamomeus.  
 OS  
 XX WO200129187-A1.  
 PN 26-APR-2001.  
 PD  
 PF 13-OCT-2000; 2000WO-JP07135.  
 XX 18-OCT-1999; 99JP-0295649.  
 PR (AJIN ) AJINOMOTO CO INC.  
 PA Taguchi S, Momose H;  
 PI WPI; 2001-300330/31.  
 DR N-PSDB; AAB20187.  
 XX Streptomyces sp. carrying an actinomycete-derived gene and promoter for  
 PT producing high yields of transglutaminase  
 PS Claim 6; Page 26-28; 41pp; Japanese.  
 XX The present invention describes a Streptomyces sp. containing a gene  
 CC construct comprising actinomycete-derived transglutaminase gene and  
 CC promoter. Also described are methods for producing pro-transglutaminase  
 CC and active transglutaminase. The gene construct can be used in the  
 CC production of large amounts of transglutaminase. The present sequence  
 CC represents Streptovorticillum cinnamomeum IFO 12852 transglutaminase,  
 CC which is used in the present invention.  
 XX Sequence 416 AA;  
 SQ  
 Query Match 82.1%; Score 1487; DB 22; Length 416;  
 Best Local Similarity 81.5%; Pred. NO. 2.6e-124;  
 Matches 269; Conservative 26; Mismatches 35; Indels 0; Gaps 0;  
 QY 2 SDDRVTTPAEPDLPDPYRPSYGRAETVNNYIRKWOQVYSHRDGRKQOMTEQREWLS 61  
 DB 87 sddretppaepldrmpayrayaggrattvnnnyirkwqvyshrdgkqgkqteeqrekl 146  
 QY 62 YGCVGVTVWNSGOYPTNRLAFASFDEDFRKNELKNGRPRSGETRAEFEGRVAKESFDEEK 121  
 DB 147 ygcvgvtywvnsqgpyptnrlafasfdenkyndkntsprdpdtraefegriakgsfdegk 206  
 QY 122 GFORAREVASVMNRALENAHDEGSAYLDNLKELANGNDALRNEDARSFPYSALRNTSPSK 181  
 DB 207 gfkardvasvmnkalenahdegtyinnlktelnnndallredsrnsfysalrntpsfk 266  
 QY 182 ERNGGNHDPGRMKAVIYSKHFWSGQDRSSADRKRYGDPDAFRPAGTGLVDMSSDRNP 241  
 DB 267 erdgngnydpkrmkavlyskhfwsgdrgsgskrykgydpeafpddggtglvdmksdrsp 326  
 QY 242 RSPTSPGCEGVNFDYGFAGQATEADAKTVTHGNHYPAPNGSLGAMHVESKFRNWSG 301  
 DB 327 rspakpgegvnfdygfagqateadaktwthghyapnsglghpmhveskfrkwsag 386  
 QY 302 YSDFDRGAYVITFIPKSWNTAPDKVKQGW 331  
 DB 301 yadfdrgayvitfipkswntapakvegwp 330

Db 387 yadfrgavvifipkswntapakveggwp 416

Search completed: September 27, 2002, 12:39:35  
Job time: 88 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2002, 12:38:07 ; Search time 13.02 Seconds  
(without alignments)  
620.958 Million cell updates/sec

Title: US-09-884-948-1

Perfect score: 1811  
Sequence: 1 DSDRVPAPAEPLDMPDY.....ITFIPKSWNTAPDKVKQGW 331

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1811	100.0	331	1	US-08-136-993-1
2	1811	100.0	331	3	US-09-109-063-1
3	1811	100.0	331	3	US-08-793-426A-3
4	1811	100.0	331	4	US-09-294-565-3
5	1811	100.0	406	1	US-08-136-993-13
6	1465	80.9	331	3	US-08-793-426A-2
7	1465	80.9	331	4	US-09-294-565-2
8	591	32.6	126	3	US-08-793-426A-7
9	591	32.6	126	4	US-09-294-565-7
10	208	11.5	40	3	US-08-793-426A-8
11	208	11.5	40	4	US-09-294-565-8
12	189	10.4	52	3	US-08-793-426A-6
13	189	10.4	52	4	US-09-294-565-6
14	99	5.5	315	3	US-08-558-135-5
15	91	5.0	544	1	US-08-264-002-7
16	90.5	5.0	357	3	US-09-041-889-30
17	87.5	4.8	1128	1	US-08-111-939-2
18	86.5	4.8	913	1	US-08-487-890A-5
19	86.5	4.8	913	2	US-08-478-435-5
20	86.5	4.8	913	2	US-08-337-483-5
21	86.5	4.8	913	2	US-08-478-373-5
22	86.5	4.8	913	3	US-08-474-671-5
23	86.5	4.8	913	3	US-08-483-577A-5
24	86.5	4.8	913	4	US-08-597-438-5
25	86.5	4.8	913	4	US-08-637-654-5
26	84.5	4.7	625	4	US-09-360-197-15
27	84.5	4.7	674	3	US-08-893-852A-1

28 84 4.6 453 4 US-08-961-083-38 Sequence 38, Appl  
29 84 4.6 535 3 US-08-725-459B-18 Sequence 18, Appl  
30 84 4.6 630 3 US-08-725-459B-17 Sequence 17, Appl  
31 84 4.6 680 3 US-08-725-459B-3 Sequence 3, Appl  
32 84 4.6 730 3 US-08-725-459B-2 Sequence 2, Appl  
33 84 4.6 733 3 US-08-725-459B-23 Sequence 23, Appl  
34 84 4.6 733 3 US-08-725-459B-28 Sequence 28, Appl  
35 84 4.6 733 3 US-08-725-459B-29 Sequence 29, Appl  
36 84 4.6 733 3 US-08-725-459B-30 Sequence 30, Appl  
37 84 4.6 748 3 US-08-725-459B-24 Sequence 24, Appl  
38 84 4.6 748 3 US-08-725-459B-27 Sequence 27, Appl  
39 84 4.6 753 3 US-08-725-459B-25 Sequence 25, Appl  
40 84 4.6 753 3 US-08-725-459B-26 Sequence 26, Appl  
41 84 4.6 769 3 US-08-725-459B-35 Sequence 35, Appl  
42 84 4.6 769 3 US-08-725-459B-36 Sequence 36, Appl  
43 84 4.6 769 3 US-08-725-459B-37 Sequence 37, Appl  
44 84 4.6 769 3 US-08-725-459B-38 Sequence 38, Appl  
45 84 4.6 769 3 US-08-725-459B-39 Sequence 39, Appl

#### ALIGNMENTS

RESULT 1  
US-08-136-993-1  
; Sequence 1, Application US/08136993  
; Patent No. 5420025  
; GENERAL INFORMATION:  
; APPLICANT: Takagi, Hiroshi  
; APPLICANT: Arafuka, Shino  
; APPLICANT: Matsui, Hiroshi  
; APPLICANT: Washizu, Kinya  
; APPLICANT: Ando, Keiichi  
; APPLICANT: Koikeda, Satoshi  
; TITLE OF INVENTION: Recombinant transglutaminase  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
; STREET: 2100 Pennsylvania Avenue  
; CITY: N.W.  
; STATE: Washington, D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20037-3202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/136,993  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/777,447  
; FILING DATE:  
; APPLICATION NUMBER: JP 2-282566  
; FILING DATE: 19-OCT-1990  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-293-7060  
; TELEFAX: 202-293-7860  
; TELEX: 6491103  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 331 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-136-993-1  
Query Match 100.0%; Score 1811; DB 1; Length 331;  
Best Local Similarity 100.0%; Pred. No. 3.8e-171;  
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDRMPDPYRPSYGRAETVNNYIRKWOQVYSHRDGRKQKQMTTEQREWL 60  
Db 1 DSDDRVTPPAEPLDRMPDPYRPSYGRAETVNNYIRKWOQVYSHRDGRKQKQMTTEQREWL 60  
QY 61 SYGCVGVTWVNSQYPTNRLAFASDEDFRKNELKNGRPSGTRAEFEGRVAKESFDEE 120  
Db 61 SYGCVGVTWVNSQYPTNRLAFASDEDFRKNELKNGRPSGTRAEFEGRVAKESFDEE 120  
QY 121 KGFQREVASVNNRALENNAHDESAVLDNLKELANGNDALRNEDARSFFYSALRNTPSF 180  
Db 121 KGFQREVASVNNRALENNAHDESAVLDNLKELANGNDALRNEDARSFFYSALRNTPSF 180  
QY 181 KERNGNHDSRMKAVIYKSHFWSGQDRSSADKRYGDPDAFRPAPGTGLVDMSDRNI 240  
Db 181 KERNGNHDSRMKAVIYKSHFWSGQDRSSADKRYGDPDAFRPAPGTGLVDMSDRNI 240  
QY 241 PRSPTSPGEGFVNFYDYGWGAQTEADADKTVWTHGNHYHAPNGSLGAMHYVESKFRNWSE 300  
Db 241 PRSPTSPGEGFVNFYDYGWGAQTEADADKTVWTHGNHYHAPNGSLGAMHYVESKFRNWSE 300  
QY 301 GYSDFDRGAYVITFIPKSWNTAPDKVKQGW 331  
Db 301 GYSDFDRGAYVITFIPKSWNTAPDKVKQGW 331

## RESULT 2

US-09-109-063-1  
; Sequence 1, Application US/09109063  
; Patent No. 6013498  
; GENERAL INFORMATION:  
; APPLICANT: YOKOYAMA, KEIICHI  
; APPLICANT: NAKAMURA, NAMI  
; APPLICANT: MIWA, TETSUYA  
; APPLICANT: SEGURO, KATSUYA  
; TITLE OF INVENTION: PROCESS FOR PRODUCING MICROBIAL TRANSGLUTAMINASE  
; FILE REFERENCE: 0010-0937-0  
; CURRENT APPLICATION NUMBER: US/09/109,063  
; CURRENT FILING DATE: 1998-07-02  
; EARLIER APPLICATION NUMBER: JP 180010/1997  
; EARLIER FILING DATE: 1997-07-04  
; NUMBER OF SEQ ID NOS: 62  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 331  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial  
; OTHER INFORMATION: Sequence: TRANSGLUTAMINASE  
US-09-109-063-1

Query Match 100.0%; Score 1811; DB 3; Length 331;  
Best Local Similarity 100.0%; Pred. No. 3, 8e-171;  
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDRMPDPYRPSYGRAETVNNYIRKWOQVYSHRDGRKQKQMTTEQREWL 60  
Db 1 DSDDRVTPPAEPLDRMPDPYRPSYGRAETVNNYIRKWOQVYSHRDGRKQKQMTTEQREWL 60  
QY 61 SYGCVGVTWVNSQYPTNRLAFASDEDFRKNELKNGRPSGTRAEFEGRVAKESFDEE 120  
Db 61 SYGCVGVTWVNSQYPTNRLAFASDEDFRKNELKNGRPSGTRAEFEGRVAKESFDEE 120  
QY 121 KGFQREVASVNNRALENNAHDESAVLDNLKELANGNDALRNEDARSFFYSALRNTPSF 180  
Db 121 KGFQREVASVNNRALENNAHDESAVLDNLKELANGNDALRNEDARSFFYSALRNTPSF 180  
QY 181 KERNGNHDSRMKAVIYKSHFWSGQDRSSADKRYGDPDAFRPAPGTGLVDMSDRNI 240  
Db 181 KERNGNHDSRMKAVIYKSHFWSGQDRSSADKRYGDPDAFRPAPGTGLVDMSDRNI 240

QY 241 PRSPTSPGEGFVNFYDYGWGAQTEADADKTVWTHGNHYHAPNGSLGAMHYVESKFRNWSE 300  
Db 241 PRSPTSPGEGFVNFYDYGWGAQTEADADKTVWTHGNHYHAPNGSLGAMHYVESKFRNWSE 300  
QY 301 GYSDFDRGAYVITFIPKSWNTAPDKVKQGW 331  
Db 301 GYSDFDRGAYVITFIPKSWNTAPDKVKQGW 331  
RESULT 3  
US-08-793-426A-3  
; Sequence 3, Application US/08793426A  
; Patent No. 6100053  
; GENERAL INFORMATION:  
; APPLICANT: Bech, Lisbeth  
; APPLICANT: No. 6100053revang, Iben  
; APPLICANT: Halkier, Torben  
; APPLICANT: Rasmussen, Grethe  
; APPLICANT: Schafer, Thomas  
; APPLICANT: Andersen, Jens  
; TITLE OF INVENTION: Microbial Transglutaminases, Their  
; TITLE OF INVENTION: Production And Use  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 61000530 No. 6100053disk of No. 6100053th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/793,426A  
; FILING DATE: 25-FEB-1997  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Rozek, Carol E.  
; REGISTRATION NUMBER: 36,993  
; REFERENCE/DOCKET NUMBER: 4211.204-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 331 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-793-426A-3

Query Match 100.0%; Score 1811; DB 3; Length 331;  
Best Local Similarity 100.0%; Pred. No. 3, 8e-171;  
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTPPAEPLDRMPDPYRPSYGRAETVNNYIRKWOQVYSHRDGRKQKQMTTEQREWL 60  
Db 1 DSDDRVTPPAEPLDRMPDPYRPSYGRAETVNNYIRKWOQVYSHRDGRKQKQMTTEQREWL 60  
QY 61 SYGCVGVTWVNSQYPTNRLAFASDEDFRKNELKNGRPSGTRAEFEGRVAKESFDEE 120  
Db 61 SYGCVGVTWVNSQYPTNRLAFASDEDFRKNELKNGRPSGTRAEFEGRVAKESFDEE 120  
QY 121 KGFQREVASVNNRALENNAHDESAVLDNLKELANGNDALRNEDARSFFYSALRNTPSF 180  
Db 121 KGFQREVASVNNRALENNAHDESAVLDNLKELANGNDALRNEDARSFFYSALRNTPSF 180

QY 181 KERNGNHDPSRMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVDMSSDRNI 240  
|||||  
Db 181 KERNGNHDPSRMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVDMSSDRNI 240  
|||||  
QY 241 PRSPTSPGEGFVNFYDYGWFGAQTEADADKTVWTHGNHYHAPNGSLGAMHVYESKFRNWSE 300  
|||||  
Db 241 PRSPTSPGEGFVNFYDYGWFGAQTEADADKTVWTHGNHYHAPNGSLGAMHVYESKFRNWSE 300  
|||||  
QY 301 GYSDFDRGAYVITFIPKSWNTAPDKVKGQWP 331  
|||||  
Db 301 GYSDFDRGAYVITFIPKSWNTAPDKVKGQWP 331  
|||||  
RESULT 4  
US-09-294-565-3  
; Sequence 3, Application US/09294565  
; Patent No. 6190879  
; GENERAL INFORMATION:  
; APPLICANT: Bech, Lisbeth  
; APPLICANT: No. 6190879revang, Iben  
; APPLICANT: Halkier, Torben  
; APPLICANT: Rasmussen, Grethe  
; APPLICANT: Schafer, Thomas  
; APPLICANT: Andersen, Jens  
; TITLE OF INVENTION: Microbial Transglutaminases, Their  
; TITLE OF INVENTION: Production And Use  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 61908790 No. 6190879disk of No. 6190879th America, Inc.  
; STREET: 405 Lexington Avenue  
; CITY: New York  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 10174  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/294,565  
; FILING DATE: 19-APR-1999  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Green, Reza  
; REGISTRATION NUMBER: 38,475  
; REFERENCE/DOCKET NUMBER: 4211.224-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 212-867-0123  
; TELEFAX: 212-878-9655  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 331 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-294-565-3

Query Match 100.0%; Score 1811; DB 4; Length 331;  
Best Local Similarity 100.0%; Pred. No. 3,8e-171;  
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTTPAEPDLRMPDPYRPSYGRAETVNNYIRKQVYSHRDGRKQOMTEQREWL 60  
|||||  
Db 1 DSDDRVTTPAEPDLRMPDPYRPSYGRAETVNNYIRKQVYSHRDGRKQOMTEQREWL 60  
|||||  
QY 61 SYGCGVTVWNSGQVPTNRLAFASDEDRFKNELKNGRPSRGETRAEFEGRVAKESFDEE 120  
|||||  
Db 61 SYGCGVTVWNSGQVPTNRLAFASDEDRFKNELKNGRPSRGETRAEFEGRVAKESFDEE 120  
|||||

QY 121 KGFQREVASVNNRALENNAHDSAYLDNLKKELANGNDALRNEDARSPPYSALRNTPSF 180  
|||||  
Db 121 KGFQREVASVNNRALENNAHDSAYLDNLKKELANGNDALRNEDARSPPYSALRNTPSF 180  
|||||  
QY 181 KERNGNHDPSRMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVDMSSDRNI 240  
|||||  
Db 181 KERNGNHDPSRMKAVIYSKHFWSGQDRSSADKRYGDPDAFRPAPGTGLVDMSSDRNI 240  
|||||  
QY 241 PRSPTSPGEGFVNFYDYGWFGAQTEADADKTVWTHGNHYHAPNGSLGAMHVYESKFRNWSE 300  
|||||  
Db 241 PRSPTSPGEGFVNFYDYGWFGAQTEADADKTVWTHGNHYHAPNGSLGAMHVYESKFRNWSE 300  
|||||  
QY 301 GYSDFDRGAYVITFIPKSWNTAPDKVKGQWP 331  
|||||  
Db 301 GYSDFDRGAYVITFIPKSWNTAPDKVKGQWP 331  
|||||  
RESULT 5  
US-08-136-993-13  
; Sequence 13, Application US/08136993  
; Patent No. 5420025  
; GENERAL INFORMATION:  
; APPLICANT: Takagi, Hiroshi  
; APPLICANT: Arafuka, Shinobu  
; APPLICANT: Matsui, Hiroshi  
; APPLICANT: Washizu, Kinya  
; APPLICANT: Ando, Keiichi  
; APPLICANT: Koikeda, Satoshi  
; TITLE OF INVENTION: Recombinant transglutaminase  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sughrue, Mion, Zinn, Macpeak & Seas  
; STREET: 2100 Pennsylvania Avenue  
; CITY: N.W.  
; STATE: Washington, D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20037-3202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/136,993  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/777,447  
; FILING DATE:  
; APPLICATION NUMBER: JP 2-282566  
; FILING DATE: 19-OCT-1990  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-293-7060  
; TELEFAX: 202-293-7860  
; TELEX: 6491103  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 406 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-136-993-13

Query Match 100.0%; Score 1811; DB 1; Length 406;  
Best Local Similarity 100.0%; Pred. No. 5.1e-171;  
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 DSDDRVTTPAEPDLRMPDPYRPSYGRAETVNNYIRKQVYSHRDGRKQOMTEQREWL 60  
|||||  
Db 76 DSDDRVTTPAEPDLRMPDPYRPSYGRAETVNNYIRKQVYSHRDGRKQOMTEQREWL 135  
|||||  
QY 61 SYGCGVTVWNSGQVPTNRLAFASDEDRFKNELKNGRPSRGETRAEFEGRVAKESFDEE 120  
|||||

```

Db   136 SYGCVGVTVNWSQYPTNRLAFASDEDFRKNELKNGRPASGETRAEFGRVAKESFDEE 195
      |||
Qy   121 KGFQREAVSYVMNRALENAHDESAYLDNLKKLANGNDALRNEDARSFPYSALRNTPSF 180
      |||
Db   196 KGFQREAVSYVMNRALENAHDESAYLDNLKKLANGNDALRNEDARSFPYSALRNTPSF 255
      |||
Qy   181 KERNGNHDPSPMKAVIYSKHFWSCQDRSSADRRKKYGDDPAFPAGCTGLVDMSRDRI 240
      |||
Db   256 KERNGNHDPSPMKAVIYSKHFWSCQDRSSADRRKKYGDDPAFPAGCTGLVDMSRDRI 315
      |||
Qy   241 PRSFTSPGEGFYNFIDYGFAGCTEADADAKTYWTHGNHYHAPNGSLGAMHYVESKFRNWSE 300
      |||
Db   316 PRSFTSPGEGFYNFIDYGFAGCTEADADAKTYWTHGNHYHAPNGSLGAMHYVESKFRNWSE 375
      |||
Qy   301 GYSDFDRGAYVTTFPKSWNTAPDKVKOGWP 331
      |||
Db   376 GYSDFDRGAYVTTFPKSWNTAPDKVKOGWP 406
      |||

RESULT        6
US-08-793-426A-2
; Sequence 2, Application US/08793426A
; Patent No. 6100053
; GENERAL INFORMATION:
; APPLICANT: Bech, Lisbeth
; APPLICANT: No. 6100053revang, Iben
; APPLICANT: Haikier, Torben
; APPLICANT: Rasmussen, Grethe
; APPLICANT: Schafer, Thomas
; APPLICANT: Andersen, Jens
; TITLE OF INVENTION: Microbial Transglutaminases, Their
;     PRODUCTION AND USE
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6100053o No. 6100053disk of No. 6100053th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793.426A
; FILING DATE: 25-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol E.
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 4211.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 331 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-793-426A-2

Query Match          80.9%; Score 1465; DB 3; Length 331;
Best Local Similarity 79.3%; Pred. No. 6.5e-137;
Matches 261; Conservative 34; Mismatches 34; Indels 0; Gaps 0;

Qy   2 SDDRVTTPAEPLDMPDPYRPSGRAETVVNNYIRKWOQVSHRGKRQQMTEGEQEWLS 61
      |||

```

```
QY 2 SDEVTTPAELDRMPDPYPSYGRAETVYNNYIRKQOYSHRDGRKQOMTEOEOWLS 61
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 YGCVGTWVNSGOPYNRLAFASFEDEDFKNEKNGRPSRGETRABFEGRVAKESFDEEK 121
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 122 GFORAREVASVMNRALENADHESAYLDNLKELANGNDALRNEADSPYSALRNTPSFK 181
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 182 ERGGNHDPKMKAVIYSHFWGQDRSSADKRYKGDPAFPAFGTGLVDMKSRDRNIP 241
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 242 RSTPSCEGVNFDYGFAGQTEADADKTVWTHGNHYHAPNGSLGAMHYVESKFRNWSG 301
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 302 YSDFRGAYVITPIKSWNTAPDKVKQGW 330
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
RESULT 8
US-08-793-426A-7
; Sequence 7, Application US/08793426A
; Patent No. 6100053
; GENERAL INFORMATION:
; APPLICANT: Bech, Lisbeth
; APPLICANT: No. 6100053revang, Iben
; APPLICANT: Halkier, Torben
; APPLICANT: Rasmussen, Grethe
; APPLICANT: Schafer, Thomas
; APPLICANT: Andersen, Jens
; TITLE OF INVENTION: Microbial Transglutaminases, Their
; TITLE OF INVENTION: Production And Use
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6100053o No. 6100053disk of No. 6100053th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,426A
; FILING DATE: 25-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol E.
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 4211.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
```

```
; FRAGMENT TYPE: internal
US-08-793-426A-7
Query Match 32.6%; Score 591; DB 3; Length 126;
Best Local Similarity 82.5%; Pred. No. 4.6e-51;
Matches 104; Conservative 8; Mismatches 14; Indels 0; Gaps 0;
QY 165 DARSFEYSALRNTPEFKERNNGNDPDSRMKAVIYSHKFWGQDRSSADKRYKGDPAF 224
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 225 PAPGGLVDMKSRDRNIPRSPGEGFVNFYGFAGQTEADADKTVWTHGNHYHAPNGS 284
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 285 LGAMHV 290
Db :|||:|
Db 121 MGPMBV 126
RESULT 9
US-09-294-565-7
; Sequence 7, Application US/09294565
; Patent No. 6190879
; GENERAL INFORMATION:
; APPLICANT: Bech, Lisbeth
; APPLICANT: No. 6190879revang, Iben
; APPLICANT: Halkier, Torben
; APPLICANT: Rasmussen, Grethe
; APPLICANT: Schafer, Thomas
; APPLICANT: Andersen, Jens
; TITLE OF INVENTION: Microbial Transglutaminases, Their
; TITLE OF INVENTION: Production And Use
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6190879o No. 6190879disk of No. 6190879th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/294,565
; FILING DATE: 19-APR-1993
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4211.224-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
US-09-294-565-7
Query Match 32.6%; Score 591; DB 4; Length 126;
```



```
; APPLICANT: Andersen, Jens
; TITLE OF INVENTION: Microbial Transglutaminases, Their
; TITLE OF INVENTION: Production And Use
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 61000530 No. 6100053disk of No. 6100053th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/793,426A
; FILING DATE: 25-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol E.
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 4211.204-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; US-08-793-426A-6

Query Match 10.4%; Score 189; DB 3; Length 52;
Best Local Similarity 70.8%; Pred. No. 7.4e-12;
Matches 36; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 106 AEEGVRKESFDEKQFORAREVASYMNRALENAHDESAIDLKTELAN 156
|||||:| | :|:|||||:|:|:| | | | | | | | | | | | | | | | | |
Db 1 AEEGRIAKGXFDAFGFKRAREVASVMNKALDSAHDEGTIDHLKTELAN 51

RESULT 13
US-09-294-565-6
; Sequence 6, Application US/09294565
; Patent No. 6190879
; GENERAL INFORMATION:
; APPLICANT: Bech, Lisbeth
; APPLICANT: No. 6190879revang, Iben
; APPLICANT: Haikier, Torben
; APPLICANT: Rasmussen, Grethe
; APPLICANT: Schafer, Thomas
; APPLICANT: Andersen, Jens
; TITLE OF INVENTION: Microbial Transglutaminases, Their
; TITLE OF INVENTION: Production And Use
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6190879o No. 6190879disk of No. 6190879th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
```

```
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA: US/09/294,565
; APPLICATION NUMBER: US/09/294,565
; FILING DATE: 19-APR-1999
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Green, Reza
; REGISTRATION NUMBER: 38,475
; REFERENCE/DOCKET NUMBER: 4211.224-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 52 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; US-09-294-565-6

Query Match 10.4%; Score 189; DB 4; Length 52;
Best Local Similarity 70.6%; Pred. No. 7.4e-12;
Matches 36; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 106 AEEGVRKESFDEKQFORAREVASYMNRALENAHDESAIDLKTELAN 156
|||||:| | :|:|||||:|:|:| | | | | | | | | | | | | | | | | |
Db 1 AEEGRIAKGXFDAFGFKRAREVASVMNKALDSAHDEGTIDHLKTELAN 51

RESULT 14
US-08-558-135-5
; Sequence 5, Application US/08558135
; Patent No. 6090631
; GENERAL INFORMATION:
; APPLICANT: Catterall, William A.
; APPLICANT: Sheng, Zu-Hang
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR SCREENING
; TITLE OF INVENTION: FOR PRESYNAPTIC CALCIUM CHANNEL BLOCKERS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/558,135
; FILING DATE: 13-NOV-1995
; CLASSIFICATION: 436
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.602C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 315 amino acids
; TYPE: amino acid
; STRANDEDNESS:
```

QY 98 -----RPRSGTAEFEGRVAKE---SPDEEKGFQR-----AREVASVMNRALE 138  
Db 145 IISLLDYFVHRGPNGAHVCWVE--VLGENLLSLIQSYGHRGVPVGVKQIAIYQLLIALLD 202  
QY 139 NAHDESAYL-DNLKKE---LANGNDALRNEDARSPFYSALENTPTSPFKERNGNHDPGRMK 194  
Db 203 YLHRECGIITDLKPEENVLICIDQALQHEAPATTSSPTSSNTSSSKTRNTGY---TAK 259  
QY 195 AVIYSKHFWSGDRSSSADKREKYGDPAFRPAPGTGLV----- 232  
Db 260 APIIKR-----GQSDVNSAQERKTKAKNPTKNSKPAGQVIPSSTSTLSRFPSPLEGAYSE 315  
QY 233 ----DMSDRNIPRSPSPGEGFVNFYGNWFGAOTE-----ADADKTVWTHGN---- 276  
Db 316 ISLRDSQKHNSHPNSPSSGDSNLSILD-GVNGSOEPVPKITVKIADLGNACWTRKHFND 374  
QY 277 ----HYHAPNGSLGAMHVYESKERNWSEGYSDFD--RGAYVITEIPKSWNT 321  
Db 375 VOTROYRSPVILGCR--WGASADCWSPACIIFELLTGDL--FDPNGNS 421

Search completed: September 27, 2002, 12:38:29  
Job time: 22 sec

QY 99 PRSGTAEFEGRVAKESSDEEKGFORAREVASVMNRALENAHDESAYLDNLKELANGN 158  
Db 107 PQENRNNNTKSVABPTVDORLGOQRAEDFLKQARHHRADRSAAH----- 154  
QY 159 DALRNEDARSPFYSA-----LRNTPSFE-----RNGGNHDPGRMKAVIYSKHFWSGQDR 208  
Db 155 -AAAGLDARPWAGSOEAEISREGPYGRESHQAREGGLEPPG-----FWEGE-- 201  
QY 209 SSSADKREKYGDPAFRPAPGTGLVDMNRDN-IPRSPSTPGE 249  
Db 202 --AERKAGDPRH-RHAHQGVGGSGRSRSPRGTADGE 239

RESULT 15  
US-08-264-002-7  
; Sequence 7, Application US/08264002  
; Patent No. 5559019  
; GENERAL INFORMATION:  
; APPLICANT: GUI, JIAN-FANG  
; APPLICANT: FU, XIANG-DONG  
; TITLE OF INVENTION: NOVEL PROTEIN SERINE KINASE, SRPK1  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SPENSLEY HORN JUBAS & LUBITZ  
; STREET: 1880 Century Park East, Fifth Floor  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90067  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: US/08/264,002  
; APPLICATION NUMBER: US/08/264,002  
; FILING DATE: 22-JUN-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: TUMARKIN PH.D., LISA A.  
; REGISTRATION NUMBER: P-38,347  
; REFERENCE/DOCKET NUMBER: PD3590  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 619/455-5100  
; TELEFAX: 619/455-5110  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 544 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-264-002-7

Query Match 5.0%; Score 91; DB 1; Length 544;  
Best Local Similarity 20.7%; Pred. No. 1.2;  
Matches 85; Conservative 48; Mismatches 148; Indels 130; Gaps 18;  
QY 27 AETVNNVIRKWOQVYSHRDGR-----KQMTREQRLSYGCVGVTVWNS----- 72  
Db 25 SEKVITGNKNVKNSEVDGKSMVEKVTHEENAEHYGGHPVYIGEEFHHRYVVE 84  
QY 73 -----GOYPTNRLAF-----ASFDEDRFKNELKNG----- 97  
Db 85 RKLGGHFSTVWLAYDRAAKRRYALKVVRSAAEHYRETISIDEIRLQKIREGDEKHLGKH 144

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2002, 12:38:07 ; Search time 19.4 seconds  
(without alignments)  
1639.462 Million cell updates/sec

Title: US-09-884-948-1  
Perfect score: 1811  
Sequence: 1 DSDDRVTTPAEPDLRMDPPY.....ITPIPKSWNTAPDKVKGWMP 331  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1811	100.0	406	JC2089	protein-glutamine
2	1802	99.5	332	JC7310	protein-glutamine
3	114.5	6.3	1217	S53714	sericin1B - silkwo
4	106.5	5.9	918	E82486	proteinase VCA0223
5	105.5	5.8	1888	T14273	zinc finger protei
6	105	5.8	325	T38308	hypothetical prote
7	105	5.8	920	T52426	dynammin-like prote
8	105	5.8	1092	H96619	protein T30E16.17
9	104	5.7	575	S03745	beta-amyrase (EC 3
10	102	5.6	219	B90045	hypothetical prote
11	102	5.6	504	T31784	hypothetical prote
12	101.5	5.6	283	AH3457	acriflavin resista
13	100.5	5.5	926	B37271	A-alpha Y 3 protei
14	100	5.5	517	A39038	1-caldesmon, nonmu
15	99	5.5	2273	I46477	calcium channel BI
16	99	5.5	2424	I46480	calcium channel BI
17	98.5	5.4	1507	B47328	natural killer cel
18	98.5	5.4	2924	T18378	variant-specific s
19	98	5.4	350	S25525	outer membrane por
20	98	5.4	547	A55575	puff-specific nucl
21	98	5.4	1403	A47328	natural killer cel
22	98	5.4	3938	T42761	Bassoon protein -
23	97.5	5.4	625	A34615	profilaggrin - rat
24	97	5.4	572	S18732	autoantigen, 64K -
25	97	5.4	591	D89783	RGD-containing lip
26	97	5.4	1097	T31504	hypothetical prote
27	96.5	5.3	548	T28910	hypothetical prote
28	96	5.3	350	AF0543	outer membrane por
29	96	5.3	992	B86237	protein F14N23.17

ALIGNMENTS

RESULT 1

JC2089  
protein-glutamine gamma-glutamyltransferase (EC 2.3.2.13) precursor - Streptovorticil  
N: Alternate names: glutamyl-peptide--amine gamma-glutamyltransferase; transglutamin  
C: Species: Streptovorticillum sp.  
C: Date: 14-Jul-1994 #sequence revision 14-Jul-1994 #text\_change 07-May-1999  
C: Accession: JC2089; JC2090; A46730  
R: Washizu, K.; Ando, K.; Koikeda, S.; Hirose, S.; Matsuura, A.; Takagi, H.; Motoki, M  
Biosci. Biotechnol. Biochem. 58, 82-87, 1994  
A: Title: Molecular cloning of the gene for microbial transglutaminase from Streptover  
A: Reference number: JC2089; MUID: 94162748  
A: Accession: JC2089  
A: Molecule type: DNA  
A: Residues: 1-406 <WAS>  
A: Experimental source: strain S-8112  
R: Takehana, S.; Washizu, K.; Ando, K.; Koikeda, S.; Takeuchi, K.; Matsui, H.; Motoki,  
Biosci. Biotechnol. Biochem. 58, 88-92, 1994  
A: Title: Chemical synthesis of the gene for microbial transglutaminase from Streptove  
A: Reference number: JC2090; MUID: 94162749  
A: Accession: JC2090  
A: Molecule type: DNA  
A: Residues: 76-406 <TAK>  
R: Kanaji, T.; Ozaki, H.; Takao, T.; Kawajiri, H.; Ide, H.; Motoki, M.; Shimonishi, Y.  
J. Biol. Chem. 268, 11565-11572, 1993  
A: Title: Primary structure of microbial transglutaminase from Streptovorticillum sp.  
A: Reference number: A46730; MUID: 93280110  
A: Accession: A46730  
A: Status: preliminary  
A: Molecule type: protein  
A: Residues: 76-406 <KAN>  
A: Experimental source: s-8112  
A: Note: sequence extracted from NCBI backbone (NCBIP:133222)  
C: Comment: This enzyme catalyzes an acyl transfer reaction between a gamma-carboxyami  
lar crosslinking of certain proteins by gamma-glutamyl-epsilon-lysine side chain brid  
C: Superfamily: protein-glutamine gamma-glutamyltransferase  
C: Keywords: aminocyclotransferase; calcium; coagulation; heterotetramer; homodimer  
F: 1-18/Domain: signal sequence #status predicted <Sig>  
F: 19-75/Domain: propeptide #status predicted <PRO>  
F: 76-406/Product: protein-glutamine gamma-glutamyltransferase #status predicted <MAT>

Query Match 100.0%; Score 1811; DB 2; Length 406;  
Best Local Similarity 100.0%; Pred. No. 1.9e-128; Indels 0; Gaps 0;  
Matches 331; Conservative 0; Mismatches 0;  
QY 1 DSDDRVTTPAEPDLRMDPPYRPSYGRAETVYNNYTKRQWQYVSHRDGRKQOMTEQREWL 60  
Db 76 DSDDRVTTPAEPDLRMDPPYRPSYGRAETVYNNYTKRQWQYVSHRDGRKQOMTEQREWL 135  
QY 61 SYGCVGVTVWNSGQYPTNRLAFASDFEDFRKXELKNGRPSRGETRAEFGVAKESDEE 120  
Db 136 SYGCVGVTVWNSGQYPTNRLAFASDFEDFRKXELKNGRPSRGETRAEFGVAKESFDEE 195

QY 121 KGQREAVASYMNRALENAHDSEAYLDNLKKELANGNDALRNEDARSPYSALRNTPSF 180  
|||||  
Db 196 KGQREAVASYMNRALENAHDSEAYLDNLKKELANGNDALRNEDARSPYSALRNTPSF 255  
|||||  
QY 181 KERNGNHDPSSMKAVIYSKHFWSQDSSADKKRYGDPDAFRPAPCTGLVDMRSRDRNI 240  
|||||  
Db 256 KERNGNHDPSSMKAVIYSKHFWSQDSSADKKRYGDPDAFRPAPCTGLVDMRSRDRNI 315  
|||||  
QY 241 PRSPITSPGEGFVNDYGVWFGAQTEADADKTYVTHGNHYHAPNGSLGAMHVYESKFRNWSE 300  
|||||  
Db 316 PRSPITSPGEGFVNDYGVWFGAQTEADADKTYVTHGNHYHAPNGSLGAMHVYESKFRNWSE 375  
|||||  
QY 301 GYSDFDGRGAYVITFIPIKSWNTAPDKVKGWP 331  
|||||  
Db 376 GYSDFDGRGAYVITFIPIKSWNTAPDKVKGWP 406  
|||||

RESULT 2  
JC7310

protein-glutamine gamma-glutamyltransferase (EC 2.3.2.13) - Escherichia coli  
N Alternate names: microbial transglutaminase  
C Species: Escherichia coli  
C Date: 08-Sep-2000 #sequence\_revision 08-Sep-2000 #text\_change 17-Nov-2000  
C Accession: JC7310  
R Yokoyama, K.; Nakamura, N.; Seguro, K.; Kubota, K.  
Biosci. Biotechnol. Biochem. 64, 1263-1270, 2000  
A Title: Overproduction of microbial transglutaminase in Escherichia coli, in vitro ref  
A Reference number: JC7310  
A Accession: JC7310  
A Molecule type: DNA  
A Residues: 1-332 <YOK>  
A Experimental source: strain JM109  
A Note: Cross-reference  
C Comment: This enzyme belongs to a family of enzymes that catalyzes acyl transfer between  
in the formation of epsilon-(gamma-glutamyl) lysine cross-linkages. This enzyme is invol  
C Genetics:  
A Gene: mtg  
C Superfamily: protein-glutamine gamma-glutamyltransferase  
C Keywords: aminocyltransferase

Query Match	99.5%;	Score 1802;	DB 2;	Length 332;
Best Local Similarity	99.7%;	Pred. No. 6.9e-128;		
Matches 330;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
QY	1	DSDDRVTTPAEPLDRMPDPYRPSYGRAETVYVNYIRKWOQVYSHRDGRKQMQTEQRWL	60	
DB	2	DSDDRVTTPAEPLDRMPDPYRPSYGRAETVYVNYIRKWOQVYSHRDGRKQMQTEQRWL	61	
QY	61	SYGCGVTVWVNSGQPTNRLAFASDEDFEKELNGRPSRGETRAETEGRVAKESFDEE	120	
DB	62	SYGCGVTVWVNSGQPTNRLAFASDEDFEKELNGRPSRGETRAETEGRVAKESFDEE	121	
QY	121	KGFQREAVSYVMNRALENAHDESAYLDNLKELANGNDALNEDARSFFYSALNRTPSF	180	
DB	122	KGFQREAVSYVMNRALENAHDESAYLDNLKELANGNDALNEDARSFFYSALNRTPSF	181	
QY	181	KERNGNHDPSPRMKAVIYISKHFWSGODRSSADK9KYGDPDAFRPAGTGLVDMGRDNI	240	
DB	182	KERNGNHDPSPRMKAVIYISKHFWSGODRSSADK9KYGDPDAFRPAGTGLVDMGRDNI	241	
QY	241	PRSPSPGEGFVNFYDYGWFGAQTADADKTVTHGNHYPANGSLGAMHVYESKERNRSE	300	
DB	242	PRSPSPGEGFVNFYDYGWFGAQTADADKTVTHGNHYPANGSLGAMHVYESKERNRSE	301	
QY	301	GYSDFDORGAYVITFTPKSWNTAPDKVKQGW	331	
DB	302	GYSDFDORGAYVITFTPKSWNTAPDKVKQGW	332	

RESULT  
S52714  
3

sericin1B - silkworm  
C:Species: Bombyx mori (silkworm)  
C:Date: 19-May-1995 #sequence\_revision 21-Jul-1995 #text\_change 09-Sep-1997  
C:Accession: S52714  
R:Gareil, A.A.; Deleage, G.G.; Prudhomme, J.J.  
submitted to the EMBL Data Library, March 1995  
A:Description: Structure and organisation of the Bombyx mori sericin I gene and of th  
A:Reference number: S52714  
A:Accession: S52714  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1217 <GAR>  
A:Cross-references: EMBL:Z48802; NID:g755699; PID:g755700

	Query Match	6.38;	Score 114.5;	DB 2;	Length 1217;
	Best Local Similarity	20.4%;	Pred. No. 1.8;	136;	Indels 51; Gaps
	Matches 60;	Conservative 47;	Mismatches		
Qy	33	NYIRKWOQVYSHRDRGKRQMTEQR-	--EWLSYCGVGVTWVNSGQ-----	YPTNR	79
		:	:    :	:	:
Db	124	NVSDGQAVASSDARDENRSQAQNNAQNWADSYGYVSADRSSASRRRCQNVYSKD	183		
Qy	80	LAFASFDEDRFKNELNCRPRSGETRAFFEGRVAKESFEDEKGQRAREVASVMNRALEN	139		
		:	:    :	:	:
Db	184	ITAASKDDSRADSSRRSN-----	-AYYNRDSO-----	GSEAGLSDRSASS	223
Qy	140	AHDESAYLDNLKKELANGDNALRNEDARSPYSALRNTPFKERNGGNHDP-	----	RMK	194
		:	:    :	:	:
Db	224	SKNDNVFYRTKDSI--GGQAKSRRSSHQSSEDAYNSSPDGYNAGTRDSTSNKKKAS	281		
Qy	195	AVIYS--KHFWSGDGDRSSADKRYKGDPAFRPAPGTGLVDMSDRDNI	PRSP	TSPGEGV	252
		:	:    :	:	:
Db	282	STIIADKDIIRAANDRSSKOLKOSSAQISGP-KGTSV--SSKDRQVNDKRKSDDAYV	338		
Qy	253	NFDYGWFQ--AQTADADKTVMTHGNHYHAPNGSLGAMHVYESKFRNWSEGYSD	304		
Db	339	GRD----CTVAYSXNKDSKTSRQSNNTYADONSVRSDSAADQTSKYDRGYSD	388		

```

RESULT      4
Proteinase  VCA0223 [imported] - Vibrio cholerae (strain N16961 serogroup O1)
E82486
C:Species:  Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: E82486
R:Heidelberg, J.F.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.;
  chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
  1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: E82486
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 11918 <HEI>
A:Cross-references: GB:AE004362; GB:AE003853; NID:99657611; PIDN:AAF96135.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0223
A:Map position: 2
C:Superfamily: Bacillus thuringiensis immune inhibitor A

```

Query Match	5.9%	Score 106.5	DB 2	Length 918
Best Local Similarity	21.2%	Pred. No. 5.1		
Matches 65	Conservative 52	Mismatches 113	Indels 77	Gaps 16
QY 42	YSHR-DGRRKQQTPEQREWLSCYGVGVTVWNSGOYPTNRL-AFASFDED-RFKNELKNGR	98		
	:    :	:     :	:	:
Db 446	YSNRGDDLLKNMRP-----LTIAPGQATLRFKAWFQIEKXDYARVYLNGK	493		
QY 99	PRSGTETRA---EFGKRVAKESDFDEKFGQAR-EVAVSYNNRALENAHD-----ESA	145		

Db 494 PIAGNITMDDPKSKGPAISQSGWYDAQFDLSAWAGQTVLAFADLYTDGGLAMEGL 553  
QY 146 YLDNLKELANGDALNEDARSPPYSALNTPSK--PRNGNHDPSPMKAVIYSKHFW 203  
Db 554 YVDLLEVDNGOTLIDNAEGTSSF-----AFQFTKNGGFHE-----ANHYY 596  
QY 204 SGDRSSS-----ADKRYGDPDAPRPAPGTGLVDMRDNRIPRSPTPSGGFGVNFEDY 257  
Db 597 LLQWRSHNDVQGLANKRFGQLMSEPLLWYVDESADNWGK--HPGE-----G 647  
QY 258 WFGAQTAEADAKTVTHGHNYHAPNGSLGAMHVESKFRNWSGSDFRGAYVTFIPK 317  
Db 648 WLGV-VDAQNALVWSKTG-----EVAQTRFQVRDATPSLFDAQP--LKLVT 692  
QY 318 SWNTAPD 324  
Db 693 DGNLTED 699  
RESULT 5  
T14273  
zinc finger protein 106 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 20-Sep-1999  
C:Accession: T14273  
R:Zuberi, A.R.; Christianson, G.J.; Mendoza, L.M.; Shastri, N.; Roopenian, D.C.  
Immunity 9, 687-698, 1998  
A:Title: Positional cloning and molecular characterization of an immunodominant cytotoxic  
A:Reference number: T17953; MUID:99060924  
C:Accession: T14273  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1888 <ZUB>  
A:Cross-references: EMBL:AF060245; NID:93372656; PID:93372657; PIDN:AAD04329.1  
C:Genetics:  
A:Note: zip106  
Query Match 5.8%; Score 105.5; DB 2; Length 1888;  
Best Local Similarity 19.6%; Pred. No. 15;  
Matches 68; Conservative 37; Mismatches 101; Indels 141; Gaps 17;  
QY 8 PPAEPLRMPDPYRPGYGRAETVYNNYIRKQQVYSHRDGRKQOMTEQREWLMSGCVGV 67  
Db 109 PPSNSQVNSDDRPQWRREDIPYQDRSYSQPPRRHRG-----PPQRDW----- 154  
QY 68 TWNSGQYPTNRLAFASFDDEKFNELKN-GRPR-----SGETRA----- 106  
Db 155 KWEKDG-----FNSTRKNSPPLRLNSGGPRGSSVWHKGAIRGSSWFLNHSNGGG 206  
QY 107 -----EFEGRVAKESFDEE--KGFORAREVASVMNRALENHADESAYLDNLKEL 154  
Db 207 WHSNNGMVDNNGTGRNNSWHSEGTGGF-----PSWHMNN----- 242  
QY 155 ANGDALNEDARSPYSALRNT-----PSPKERNNGNHDPSPM-----K 194  
Db 243 SNGN-----WKSSYFSTNSWNYNGPGDKFOQGRNRNPNYQMEDTKMWNKSNK 291  
QY 195 AVIYSKH--FWSGQDRSSSADKKRYGDPDAPRPAPGTGLVDMRDNRIPRSPTPSGEGEV 252  
Db 292 PSKYSQERCKWQRDRKAAKYR--SPPEGY-----ASDTFPSEGLL 331  
QY 253 NFDYGNFGAQTAEADAKTVTHGHNYHAPNGSLGAMHVESKFRNWS 299  
Db 332 EFNF-----EQRESQTKQTDAAKINGKNGTKA-----RDKFPRWT 369  
RESULT 6  
T38308  
hypothetical protein SPAC23H3.15c - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 20-Oct-2000

C:Accession: T38308; T38381  
R:Skellton, J.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.; Wood, V.  
submitted to the EMBL Data Library, September 1997  
A:Reference number: 221785  
A:Accession: T38308  
A:Molecule type: DNA  
A:Residues: 1-325 <SKE>  
A:Cross-references: EMBL:Z99163; PIDN:CAB1245.1; GSPDB:GN00066; SPDB:SPAC23H3.15c  
A:Experimental source: strain 972h-; cosmid c23H3  
R:Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
submitted to the EMBL Data Library, August 1997  
A:Reference number: 221737  
A:Accession: T38381  
A:Molecule type: DNA  
A:Residues: 1-87 <DEV>  
A:Cross-references: EMBL:Z98944; PIDN:CAB11598.1; GSPDB:GN00066; SPDB:SPAC25H1.01c  
A:Experimental source: strain 972h-; cosmid c25H1  
C:Genetics:  
A:Gene: SPAC23H3.15c; SPDB:SPAC25H1.01c  
A:Map position: 1

Query Match 5.8%; Score 105; DB 2; Length 325;  
Best Local Similarity 23.0%; Pred. No. 1.7;  
Matches 64; Conservative 33; Mismatches 85; Indels 96; Gaps 16;  
QY 71 NSGQYPTNRLAFASFDDEKFNK-----ELKNRPRSGETRAEFEGRVAKESFDEE 120  
Db 32 SSSGAPQRNFDTSYTSEIPNSRAANDMTDGGSDPYAGMTS-----DTR 79  
QY 121 KGF-----ORAREVASVM--NRALENHADESAYLDNLKELANGNDAL----RNEDARSPE 170  
Db 80 KGFNSVESRKKKEQSDVRGGDTYSRRHDDSSYSN--KYSTGGNDSYSSGGRNED----- 132  
QY 171 YSALNTSFVERNGGNH--DPSRM-KAVIYSKHFW-----GQDRSSSADKKR 216  
Db 133 YST-----SGSYTTDPSTRTDTSYGCQSYNOSRXTTGGDYGEDYSQSYPTDT 182  
QY 217 YGDPDAFRPAPGTGLVDMRDNRIPRSPTPSGEGFVNFYGMFGAQTAEADAKTVTHGN 276  
Db 183 YG-----SRQKATPSDVTGGGA---YDSSSGSHTHGSHGTEHRGS 222  
QY 277 H--YHAPNGSLGAMHVESKFRNWSGYS--DFDRGAY 310  
Db 223 YGNDNTANKTRGAVS-----SAGYSGEGYKGTY 251  
RESULT 7  
T52426  
dynamitin-like protein [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 24-Oct-2000 #sequence\_revision 24-Oct-2000 #text\_change 24-Oct-2000  
C:Accession: T52426  
R:Kato, A.; Suzuki, M.; Kuwahara, A.; Ooe, H.; Higano-Inaba, K.; Kameda, Y.  
Gene 239, 309-316, 1999  
A:Title: Isolation and analysis of cDNA within a 300 kb Arabidopsis thaliana genomic  
A:Reference number: 225171  
A:Accession: T52426  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-920 <KAT>  
A:Cross-references: EMBL:AB028467; PIDN:BA088111.1  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Gene: CFI  
A:Map position: 1

Query Match 5.8%; Score 105; DB 2; Length 920;  
Best Local Similarity 24.6%; Pred. No. 6.6;  
Matches 59; Conservative 27; Mismatches 104; Indels 50; Gaps 11;  
QY 26 RAETVV--NNYIRKQQVYSHRDG-----RKQOMTEQREWLMSG 63

Db 634 KAESWPKNWKINLKQVIGARGQGVGASMRGSLSGSLDKMKVRKVPDPEELRWMSQE 743  
Qy 64 CVGTWVNSGOYPTN-----RLAFASDEDFRKNELANGRPMSGETRAEF---EGRVAK 115  
Db 744 VRGYVEAVLNSLAANVPKAVVLCOVERSKEDMLNQLYSSISAIGNERIESLIQEDQNVKR 803  
Qy 116 SFDEEKGFORAREVASYMNRALENAHDE----SAYLDNLKKELA---NGNDALRNEDAR 167  
Db 804 RDRD---YKQSSLLSKLTROL-SIHNRRAAASSWSDNGSTESSPTNGSS--GEDWM 857  
Qy 168 SPFYSLRNTSPFKERNGNHDPGRMKAVIYSHFWGSGDRSSADKKRYGDDPAFRPAP 227  
Db 858 NAFNAAAGPDSLXRYGSGH--SRR---YSDPAQNGEDSSGSGSRRTTPNRLPPAP 911

RESULT 8

H96619  
Protein T30E16.17 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
R:Accession: H96619  
C:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 815-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, X.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marfiali,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: AB6141; MUID:21016719  
A:Accession: H96619  
\*A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1092 <STO>  
A:Cross-references: GB:AB005173; NID:g8778745; PIDN:AAF79753.1; GSPDB:GN00141  
C:Genetics:  
\*A:Gene: T30E16.17  
A:Map position: 1

```

Query Match          5.8%; Score 105; DB 2; Length 1092;
Best Local Similarity 24.6%; Pred. No. 8.2;
Matches 59; Conservative 27; Mismatches 104; Indels 50; Gaps 11;

QY 26 RAETVY--NNVIRKQWQVYSHRDG-----RKQMTTEQREWLWSYG 63
      :|::| | :| | | | | | | | | | | | | | | | | | | | | | |
Db 856 KAESVMVDKNEWINKIQVIAQGQGVGSASMRQSLSEGLDKMYRKVPYDPEELRWMSQE 915
      :|::| | :| | | | | | | | | | | | | | | | | | | | | | |
QY 64 CVGVTVWNSGQVPTN-----RLAFASFDEDFKELNKLNGRPSRGETRAEF--EGRVAKE 115
      :|::| | :| | | | | | | | | | | | | | | | | | | | | | |
Db 916 VRGYEAVNLSLAANVPKAVLVCQVEKEDMLKOLYSSIAIGNERIESLIQEDQNVKR 975
      :|::| | :| | | | | | | | | | | | | | | | | | | | | | |
QY 116 SFDEKGFQRAEVASVMNRLAENAHDF-----SAYLDNLKELA---NGMDALRNEDAR 167
      :|::| | :| | | | | | | | | | | | | | | | | | | | | | |
Db 976 RDRD---YQKOSLSLSKLTROL-STHDNRAAAASWSNSGTSESPRTNGGSS--GEDWM 1029
      :|::| | :| | | | | | | | | | | | | | | | | | | | | | |
QY 168 SFYSGALRNTSFKERNGNHDPDSMKKAVIYSKHFWSQDRSSSADKRYKGDPAFRAP 227
      :|::| | :| | | | | | | | | | | | | | | | | | | | | | |
Db 1030 NAFNAAASGPDLSKRYGSGGH--SRR---YSDPAQNGEDSSGGSGSRRITPNRLPPAP 1083
      :|::| | :| | | | | | | | | | | | | | | | | | | | | | |

RESULT 9
S03745
beta-amylase (EC 3.2.1.2) precursor - Bacillus circulans
C:Species: Bacillus circulans
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S03745
R:Signgens, K.W.
Mol. Microbiol. 1, 86-91, 1987
A:Title: Molecular cloning and characterization of the beta-amylase gene from Bacillus

```

A:Reference number: S03745; MUID:88260890  
A:Accession: S03745  
A>Status: not compared with conceptual translation  
A:Molecule type: DNA  
A:Residues: 1-575 <SIGI>  
A:Cross-references: GB:Y00523; NID:g39407; PIDN:CAA68578.1; PID:g99408  
C:Function:  
A:Description: catalyzes the hydrolysis of 1,4-glycosidic linkages of starch, removing  
C:Superfamily: beta-amylase  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation  
F:1-36Domain: signal sequence #status predicted <SIG>  
F:37-575Product: beta-amylase #status predicted <MAT>

Query Match 5.7%; Score 104; DB 1; Length 575;  
Best Local Similarity 22.9%; Pred. No. 4.3;  
Matches 76; Conservative

Qy 24 YGRAETVNNYIRKQOVYSHRGRKQOMTEQRWLSVCVGVTWV-----NSG 73  
:  
Db 80 WGYVESAGDNQF-DWSYKYTYADVTKQ-----AGLKWPPIITHRCGGNGV 124  
:  
Qy 74 Q-----YPTNRLAFASDEDFKFKLNKGRPSQ-----ETRAEEGRVAKESEDEKGFQ 124  
:  
Db 125 DDCHNPFLSWLWKSXGADEMFKDE-----SGYVNESLSFWSGVGKQ-YDE----- 171  
:  
Qy 125 RAREVASVMNRALNAHDESAYLDNLKKELANGNDALRNEDAPSPFYSA LRNTPSFKERN 184  
:  
Db 172 ---LYASP-----AQNFSAYKDMIPKIYLSGGP-----SGELRYPSYPAA 209  
:  
Qy 185 GGNHDPPMKAVIYSKHEWSGODRSSADRKYGDPDAFRPAGTGLVDMSRDNRINPRSP 244  
:  
Db 210 GWSY-PARGKFTQVTE---TAKSFRMTATKYSLDKINNAWGNTITSMSQ-----ISP 260  
:  
Qy 245 TSPGEFGF-----VFNDYG-----WF-----GAQTADADKTWTWH-----GNH 277  
:  
Db 261 PTSDSGFTGGGINITYIKDFLSVQSVLENLHLGVAAAHKNEFDPVFGVRIGAKISGIH 320  
:  
Qy 278 YHAPNGSL--GAMHVYESKFNWSEGYSDFDR 307  
:  
Db 321 WOMNPNMPHSAAEH-----AGGIYDYNK 343  
: :

RESULT 10  
B90045  
hypothetical protein [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 22-Oct-2001  
C:Accession: B90045  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O  
ma, A.; Mizutani-Ui, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu,  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.  
Lancet 357, 1225-1240, 2001  
Article: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: AB9758; MUID:21311952; PMID:11418146  
A:Accession: B90045  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-219 <KUR>  
A:Cross-references: GB:BA000018; PID:g13702382; PIDN:BAB43523.1; GSPDB:GN00149  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA2221

```
Query Match      5.6%; Score 102; DB 2; Length 219;
Best Local Similarity 20.8%; Pred. No. 1.7;
Matches 46; Conservative 42; Mismatches 73; Indels 60; Gaps 11

QY 18 DYPKPSY-----GRAETVNNYIRKW-----QQVYSHRDGRKQOMTEQRE-- 58
          ||| | : : : : : ||| | : : : : : ||| | : : : : :
DB 5 DYKRSQYDNONHRRSDASRQOAKGDPEHPERYNGRYRREQILLEENERS 64
```

QY 59 -----WLSYGVGVWNSQYPTNRLAFASDEDRFKNELKNGRSPRSGETRAETEGRV- 112  
 Db 65 RRSKKWL-YIIIIAILIIVAFVTR--ALLNDSKVSNDPKVSNQYKKQVQENQ-DGQIN 120  
 QY 113 -----AKESFDEKGFQRAEVASVNNRALENAHDSA-----YLDNLKKELANG 157  
 Db 121 QQVDNAKENI---KNNQKTDDIIFKLQNOITDLNKQEQNKADSKLTQFYQDOINK-LTEA 176  
 QY 158 NDALRNEDARSPYSALRTPS-----FKERNGG 186  
 Db 177 NNALKNNASQKTESMLNDINTKFDISKLSLFLFADNGG 217  
 RESULT 11  
 T31784  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 31-Jan-2000  
 C:Accession: T31784  
 R:Jones, K.; Wohlmann, P.  
 submitted to the EMBL Data Library, July 1997  
 A:Description: The sequence of C. elegans cosmid F13H6.  
 A:Reference number: Z21085  
 A:Accession: T31784  
 A:Status: preliminary;  
 A:Molecule type: DNA  
 A:Map position: 5  
 A:Superfamily: cholinesterase; cholinesterase homology  
 Query Match 5.6%; Score 102; DB 2; Length 504;  
 Best Local Similarity 20.5%; Pred. No. 5.1;  
 Matches 69; Conservative 44; Mismatches 144; Indels 80; Gaps 13;  
 QY 28 ETVYNNYKRWQVYSHRGRKQMTQEQREWLSY----CVGVWNSQYPTNRLAFA 83  
 Db 149 DSVGNFGLWDTLA-----LKWQKHSSFGGDPNCVTVFGQSAGGASTLLSLUS 200  
 QY 84 SFDEDRFKNELKNGRSPRSGETRAETEGRVAK-----ESFDEKGFQRAEVASV---MN 134  
 Db 201 PHSKDLFORFI---PISGAYCEPALRTSKQAKIFREFAEFKGTGDSITLLEWYKN 256  
 QY 135 RALENAHDSAYLDNLKKELANGNDALRNEDAR----SPFYSALANTSPFKERNGGNHD 189  
 Db 257 QSSETLSD--LRKEAPKKOMTGVDEYGVIAVMNPEFSPADAGLALFPKGYGNDTAEN 314  
 QY 190 PSRMKAVIYKSHFWSGQDRSSSADKKR-----YGDPAFRPAPGTGLVDMRDRNIPRSP 244  
 Db 315 PEEMHKLFEKVV-EGVDRSDDSAMKKRLCEAFGD----- 348  
 QY 245 TSPGEGFVNFYQWF-GAOTERADADKTVTHG-NHYHAPNGSL-----GAMHVYESK 294  
 Db 349 -----LGFNLGVFSQAKSSAKYKNDVFLYSPFYHSDGFGKWKDLPLFPFASMHGTELR 401  
 QY 295 FRNWSEGYSDFRGAYVITFIPKSWNTAPDKVKQKGP 331  
 Db 402 YLLGEGFYSKFDATKEELEVEIKTITLFSNFAKYGNP 438  
 RESULT 12  
 AH3457  
 C:Species: Brucella melitensis  
 C:Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 01-Feb-2002  
 C:Accession: AH3457  
 R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,  
 ; Mazur, M.; Goltsman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letess

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit  
 A:Reference number: AD3252; PMID:1175668  
 A:Accession: AH3457  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-383 <KUR>  
 A:Cross-references: GB:AE008917; PIDN:AAL52827.1; PID:g17983667; GSPDB:GN00190  
 A:Experimental source: strain 16M  
 C:Genetics  
 A:Gene: BMEI1646  
 A:Map position: 1

Query Match 5.6%; Score 101.5; DB 2; Length 383;  
 Best Local Similarity 26.1%; Pred. No. 3.9;  
 Matches 29; Conservative 21; Mismatches 32; Indels 29; Gaps 3;  
 QY 53 GCVGVTWNSQYPTNRLAFASFD-----DRFKNELKNGRSPRSGETRAETEGR 111  
 Db 189 GIVGILPVNAGNYTAQTSIARDSTVLIDWVPERFAPIKVGQPLTAESA-FPGE 247  
 QY 112 VAKESFDEKGFQRAEVASVNNRALENAHDSAYLDNLKKELANGNDALR 162  
 Db 248 IVKGRIN-----AVDNMLDEASRTLHVRAEVPNAEDRLR 281

RESULT 13  
 B37271  
 A:alpha Y 3 protein - bracket fungus (Schizophyllum commune)  
 C:Species: Schizophyllum commune  
 C:Date: 07-Feb-1992 #sequence\_revision 07-Feb-1992 #text\_change 05-Dec-1997  
 C:Accession: B37271  
 R:Ullrich, R.C.  
 submitted to the Protein Sequence Database, October 1991  
 A:Reference number: A37271  
 A:Accession: B37271  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: DNA  
 A:Residues: 1-926 <ULL>  
 C:Superfamily: unassigned homeobox proteins; homeobox homology  
 C:Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F:148-204/Domain: homeobox homology <HOX>

Query Match 5.5%; Score 100.5; DB 2; Length 926;  
 Best Local Similarity 20.5%; Pred. No. 14;  
 Matches 71; Conservative 36; Mismatches 122; Indels 117; Gaps 15;  
 QY 9 PAEPLDRMPDYPYSYGRAETVWNNYIKWQOVYSHRGRKQMTQEQREWLSYGCYGV 68  
 Db 601 PAEPIVR-PDDFAPFVALAEKRAKRRARKEKKQAEKARK----- 640  
 QY 69 WVNSQYPTNRLAFASDEDRFKNELKNGRSPRSGETRAETEGRVAKESFDEKGFQRAE 128  
 Db 641 -----EKEARKEAKQAKDRKEQKAGLPRR-SPSLDS-----SRASS 678  
 QY 129 VASVNNRALENAHDSAYLDNLKKELANGNDALRNEDARSPFYSALRNTSPFKERNNGNH 188  
 Db 679 VTSASATSRKSRTSRKPRDSSASSVAS-----ARTP---SLSTSS--RRSGTS 724  
 QY 189 DPS--RMK---AVIYKSHFWSGQDRSSAD---KRYKGDPA----- 222  
 Db 725 MPATPRMNEPLPVVASDNFVLGTDKDVTWTPELMAQFLGEDDASGLDBPMQSEGSFSDML 784  
 QY 223 -FRPAPGTGLVDMRDRNIPRSPSPGEGFYNFYDYGWFG---AQTEADADKTVWTHGNHYH 279  
 Db 785 IFSSCNDGALGDMTADVKNPE-----LGDLSDTQLSFDMDNMWTSMDLS 828  
 QY 280 A-PNGSLGAMHVYESKFRNWSEGYSDFRGAYVITFIPKSWNTAPD 324  
 Db 829 TOPAASFDSSETSSMDPNW-----LLPQCANTAPD 859

RESULT 14  
A39038  
1-caldesmon, nonmuscle - chicken  
N:Alternate names: caldesmon, nonmuscle  
C:Species: Gallus gallus (Chicken)  
C:Date: 31-Jul-1991 #sequence\_revision 27-Jun-1994 #text\_change 22-Jun-1999  
C:Accession: A39038  
R:Hayashi K.; Fujio, Y.; Kato, I.; Sobue, K.  
J. Biol. Chem. 266, 355-361, 1991  
A:Title: Structural and functional relationships between h- and l-caldesmons.  
A:Reference number: A39038; MUID:91093148  
A:Accession: A39038  
A:Molecule type: mRNA  
A:Residues: 1-517 <HAY>  
A:Cross-references: GB:M60620; GB:M38015; NID:g212242; PIDN:AAA48936.1; PID:g212243  
A:Experimental source: brain  
C:Comment: The binding of caldesmon to F-actin is modulated by calcium and calmodulin.  
C:Comment: Two calmodulin molecules can bind to nonoverlapping domains of each caldesmon  
C:Superfamily: caldesmon  
C:Keywords: actin binding; calmodulin binding; phosphoprotein  
F:342.427.462/Binding site: phosphate (Ser) (covalent) #status predicted  
F:433.456/Binding site: phosphate (Thr) (covalent) #status predicted

```

Query Match      5.5%; Score 100; DB 1; Length 517;
Best Local Similarity 18.7%; Pred. No. 7.4;
Matches 60; Conservative 41; Mismatches 116; Indels 104; Gaps 12;

Qy 28 ETVYNKYIR-KWQGVYSHRDGRKOOMTEEGREMLSYCGVVTWVNSGYQTNRAPAFSD 86
Db 151 ETVTYSQNRNWRG-DGEDEGKKEEKSEEEK-----PREVTT-----E 188

Qy 87 EDRFK-NELKNGRPSRGETRAEFTEGRVAKESFDEBKGFQRAREVASYMNRALENA----- 140
Db 189 ENQVKDNKVKREKAPKE-EMKSVD--RKRGVPEQKAQNGERELTTPKLNSTENAFGRSN 244

Qy 141 -----HDEAYLDNLKELANGNDALNEDARSFYFYSALRNTWPSFKE 182
Db 245 LKGAANAEGSEKLKEKQQAVALDELDKRRREERRRILEEEQKKKOEAEERKIRREEE 304

Qy 183 RNGGNHDPSEMKAVIYKSHFWSGODRSSADK-----RKYGQPD 221
Db 305 KRRMKEIERRRRAAEAKRQKVPEDGVSEKKPKFCSPKGSLSUKIERAEFLNKSQKS 364

Qy 222 AFRPAPQTLGLVDMRSRDR-----NIPRSPTS-----PGEGFVANFDYGWFGAQT 263
Db 365 GMRPAHTTAVVSKIDSLREQYTSVAVGNKKAAPKAKPAASDLPYPAGVRNI----- 415

Qy 264 EADADKTVTWHGNHYHAPNGS 284
Db 416 -----KGMWEKGNVFSPPGT 431

```

```

RESULT 15
I46477
calcium channel BI-1 - rabbit
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 17-Nov-2000
C:Accession: I46477; I46478
R:Mori, Y.; Friedrich, T.; Kim, M.S.; Mikami, A.; Nakai, J.; Ruth, P.; Besse, E.; Hofman
Nature 350, 398-402, 1991
A:Title: Primary structure and functional expression from complementary DNA of a brain d
A:Reference number: I46477; MUID:91187110
A:Accession: I46477
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-2273 EMOR>
A:Cross-references: EMBL:X57476; NID:g1522; PID:CAA40714.1; PID:g1523
A:Accession: I46478
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-1856; 'H',1858, 'K',1860-1862, 'SL',1865-1866, 'VIS',1870-1876, 'K',1878-1879,

```

```

A:Cross-references: EMBL:X57698; NID:ig1524; PIDN:CA440871.1; PID:gi1525
C:Superfamily: voltage-dependent calcium channel protein alpha-1 chain

Query Match          5.5%; Score 99; DB 2; Length 2273;
Best Local Similarity 26.5%; Pred. No. 60;
Matches 43; Conservative 18; Mismatches 61; Indels 40; Gaps

Qy 99 PRSGTTRAEFTGCRVAKESFDEEKQFQARVAVSMNRALENAHDSSAYLDNLKKELANGN 158
      |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 828 POENRRNNNTINKSRVAEPTVDQLRQQQAEDFLRKARHHRDARDPSAH----- 875

Qy 159 DALRNEADSPYSYA-----LRNTPSKE-----RNGGNHDP SRMKAVIYSKHEWSQDR 208
      |||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
Db 876 -AAGLDARPWAGSQEAELSGREGYGRSDHQAEGGLEPEG-----FWEGE-- 922

Qy 209 SSSADKRYGDDPAPRPAGPTGLVDMGRDRN-TPRSPTSPGE 249
      ||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 923 ---AERGKAGDPHR-RHAHRQGVGGSGGSRGSPRTGTADGE 960

search completed: September 27, 2002, 12:38:57
Job time: 50 sec

```

Search completed: September 27, 2002, 12:38:57  
Job time: 50 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2002, 12:38:12 ; Search time 28.78 Seconds  
(without alignments)  
1989.622 Million cell updates/sec

Title: US-09-884-948-1

Perfect score: 1811

Sequence: 1 DSDRVTTPAELDRMPDPY.....ITFIPKSWNTAPDKVKQGM 331

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_todent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1811	100.0	376	Q9ZAF5	streptomyc
2	1432	79.1	411	Q9RIS1	streptomyc
3	114.5	6.3	549	Q24441	strophophila
4	114.5	6.3	889	Q9V8R8	strophophila
5	114.5	6.3	1217	Q17240	bombx mori
6	112.5	6.2	1161	Q9W2X8	strophophila
7	110	6.1	736	Q65311	adeno-asso
8	109	6.0	919	Q07351	vibrio chol
9	108	6.0	401	Q96C72	homo sapien
10	108	6.0	558	Q9H0G5	homo sapien
11	107	5.9	736	Q56139	adeno-asso
12	106.5	5.9	736	Q56137	adeno-asso
13	106.5	5.9	736	Q9WBP8	adeno-asso
14	106.5	5.9	918	Q9KMP5	vibrio chol
15	105.5	5.8	1888	Q88456	mus musculu
16	105	5.8	325	P78890	schizosacch

17	105	5.8	920	10	Q9SLT4	Q9slt4 arabidopsis
18	105	5.8	921	10	Q9SLT3	Q9slt3 arabidopsis
19	105	5.8	1092	10	Q9LO55	Q9lo55 arabidopsis
20	102.5	5.7	213	12	Q9P2V7	Q9p2v7 hepatitis d
21	102	5.6	219	16	Q99RJ9	Q99rj9 staphylococ
22	102	5.6	504	5	O16352	O16352 caenorhabdi
23	102	5.6	735	12	O56652	O56652 adeno-asso
24	100.5	5.5	467	10	Q94LN8	Q94ln8 oryza sativ
25	100	5.5	630	5	Q9GNN5	Q9gnn5 branchiosto
26	100	5.5	744	11	Q9ET77	Q9et77 mus musculu
27	100	5.5	775	5	Q9V015	Q9v015 plasmodium
28	99.5	5.5	769	5	Q95FJ0	Q95fj0 drosophila
29	99.5	5.5	939	5	Q9V9K7	Q9v9k7 drosophila
30	99	5.5	435	2	Q93QD9	Q93qd9 rhizobium m
31	99	5.5	666	4	O60592	O60592 homo sapien
32	99	5.5	746	5	Q9VH91	Q9vh91 drosophila
33	99	5.5	765	5	Q9VH92	Q9vh92 drosophila
34	99	5.5	765	5	Q9SV19	Q9sv19 drosophila
35	99	5.5	768	5	Q95NJ1	Q95nj1 drosophila
36	99	5.5	806	5	Q9SV20	Q9sv20 drosophila
37	98.5	5.4	591	2	Q9EY53	Q9ey53 staphylococ
38	98.5	5.4	2924	5	Q25733	Q25733 plasmodium
39	98	5.4	907	5	Q9X145	Q9x145 drosophila
40	98	5.4	3938	11	O88778	O88778 rattus norv
41	97.5	5.4	555	10	Q9FIE2	Q9fie2 arabidopsis
42	97.5	5.4	625	11	Q9S053	Q9s053 rattus norv
43	97	5.4	591	16	Q99X14	Q99x14 staphylococ
44	97	5.4	630	4	Q96HB5	Q96hb5 homo sapien
45	97	5.4	1097	5	Q9U2T9	Q9u2t9 caenorhabdi

#### ALIGNMENTS

RESULT 1

Q9ZAF5 ID Q9ZAF5 PRELIMINARY; PRT; 376 AA.

AC Q9ZAF5;  
DT 01-MAY-1999 (TREMBLrel. 10, Created)  
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE TRANSLUTAMINASE (EC 2.3.2.13) (FRAGMENT).  
OS Streptomyces mobaraensis.  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycetaceae; Streptomyces.  
OC NCBI\_TaxID=35621;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=DSMZ.  
RX MEDLINE=99053680; Pubmed=9839945;  
RA Pasternack R.; Dorsch S.; Otterbach J.T.; Robenek I.R.; Wolf S.,  
RA Fuchsbaue H.L.;  
RT "Bacterial pro-transglutaminase from Streptovorticillum mobaraense :  
RT purification, characterisation and sequence of the zymogen.";  
RL Eur. J. Biochem. 257:570-576(1998).  
DR EMBL; Y18315; CAA77128.1; -;  
KW Transferase; Acyltransferase.  
FT NON-ITER 1  
SQ SEQUENCE 376 AA; 42445 MW; 15FE7474DE3771B9 CRC64;

Query Match 100.0%; Score 1811; DB 2; Length 376;  
Best Local Similarity 100.0%; Pred. No. 1.4e-127;  
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DSDRVTTPAELDRMPDPYPSYGRAETVYNNYIRKQVQYSHRDGRKQKQMTTEQREWL 60  
Db 46 DSDRVTTPAELDRMPDPYPSYGRAETVYNNYIRKQVQYSHRDGRKQKQMTTEQREWL 105  
Qy 61 SYGCVGTWNSGQYPTNRLAFASFDEDRKFNELKNGRPSGETRAFEGRVAKESFDEE 120  
Db 106 SYGCVGTWNSGQYPTNRLAFASFDEDRKFNELKNGRPSGETRAFEGRVAKESFDEE 165

```

QY 121 KGFQREAVSMNRALNAHDESAVLNKKELANGNDALNEDARSPFYSALENTPSF 180
Db 166 KGFQREAVSMNRALNAHDESAVLNKKELANGNDALNEDARSPFYSALENTPSF 225
QY 181 KERNGNHDPSRMKAVIYSKHFWSGDRSSADKRYKGGDPDAFRPAGTGLVDMRDRNI 240
Db 226 KERNGNHDPSRMKAVIYSKHFWSGDRSSADKRYKGGDPDAFRPAGTGLVDMRDRNI 285
QY 241 PRSPTSPGGGFVNFYDYGWFGAOTADADKTYVTHGNHYPHAPNGSLGAMHVYESKFRNWE 300
Db 286 PRSPTSPGGGFVNFYDYGWFGAOTADADKTYVTHGNHYPHAPNGSLGAMHVYESKFRNWE 345
QY 301 GYSDFDRGAYVITFIPKSNWNTAPDKVKGWP 331
Db 346 GYSDFDRGAYVITFIPKSNWNTAPDKVKGWP 376

RESULT 2
Q9RIS1 PRELIMINARY; PRT; 411 AA.
AC Q9RIS1;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TremBLrel. 16, Last annotation update)
DE TRANSGUTAMINASE PRECURSOR (EC 2.3.2.13).
GN TGASE.
OS Streptomyces cinnamonensis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=53446;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 683.68;
RX MEDLINE=98336622; PubMed=9672751;
RA Duran R., Junqua M., Schmitter J.M., Gancet C., Goulas P.;
RT "Purification, characterization, and gene cloning of transglutaminase
RT (TGase) from Streptovorticillum cinnamonum CBS 683.68."
RL Biochimie 80:313-319(1998).
DR EMBL: Y08920; CAA70055.1;
KW Signal; Transferase; Acyltransferase.
FT SIGNAL 1 81 POTENTIAL.
FT CHAIN 82 411 POTENTIAL.
SQ SEQUENCE 411 AA; 46499 MW; 01A0CCA2EF4C388B CRC64;

```

```

Query Match 79.1%; Score 1432; DB 2; Length 411;
Best Local Similarity 79.7%; Pred. No. 3.4e-99;
Matches 263; Conservative 27; Mismatches 36; Indels 4; Gaps 3;

QY 2 SDRYTPPAELDRPDPYRPSYGRAETVNNYIRKWOQVYSHRDGRKQKQMTREQRWLS 61
Db 86 SDRYTPPAELDRPDPYRPSYGRAETVNNYIRKWOQVYSHRDGRKQKQMTREQRWLS 145
QY 62 YGCVGVTVNNSQYPTNRLAFASDFEDRFKNEKNGRPRSGETRAEFGRVAKESFDEEK 121
Db 146 YGCVGVTVNNSQYPTNRLAFASDFEDRFKNEKNGRPRSGETRAEFGRVAKESFDEEK 204
QY 122 GQREAVSMNRALNAHDESAVLNKKELANGNDALNEDARSPFYSALENTPSFK 181
Db 205 GFKRDYASVMNRALNAHDESAVLNKKELANGNDALNEDARSPFYSALENTPSFK 263
QY 182 ERNGNHDPSRMKAVIYSKHFWSGDRSSADKRYKGGDPDAFRPAGTGLVDMRDRNI 241
Db 264 ERDGNYPDKMKAVIYSKHFWSGDRSSADKRYKGGDPDAFRPAGTGLVDMRDRNI 323
QY 242 RSPTSPGGGFVNFYDYGWFGAOTADADKTYVTHGNHYPHAPNGSLGAMHVYESKFRNWE 301
Db 324 RSPAPKGGGVNFYDYGWFGAOTADADKTYVTHGNHYPHAPNGSLGAMHVYESKFRNWE 383
QY 302 YSDFDRGAYVITFIPKSNWNTAPDKVKGWP 331
Db 384 YADF--GAYVITFIPKSNWNTAPAKVEQGW 411

```

```

RESULT 3
Q24441 PRELIMINARY; PRT; 549 AA.
ID Q24441;
AC Q24441;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TremBLrel. 16, Last annotation update)
DE (CDNA2) PROTEIN 4.1 HOMOLOGUE (CORACLE) (FRAGMENT).
GN CORA OR CORACLE OR CG11949.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON R;
RX MEDLINE=94215495; PubMed=8162854;
RA Fehon R.G., Dawson I.A., Artavanis-Tsakonas S.;
RT "A Drosophila homologue of membrane-skeleton protein 4.1 is associated
RT with septate junctions and is encoded by the coracle gene."
RL Development 120:545-557(1994).
DR EMBL: L27468; AAA28742.1;
DR FlyBase; FBgn010434; cora.
KW Alternative splicing.
FT NON_TER 1
SQ SEQUENCE 549 AA; 59284 MW; A82055EF2BBA4874 CRC64;

Query Match 6.3%; Score 114.5; DB 5; Length 549;
Best Local Similarity 21.3%; Pred. No. 1.5; Indels 81; Gaps 15;
Matches 71; Conservative 42; Mismatches 42; Indels 81; Gaps 15;

QY 12 PLDRMPDPYRPSYGRAETVNNY-----IRKWOQVYSHR-----DGRKQKQMT 54
Db 68 PVDRTPEKFNRTLSGARLTGRSMDALALAEKVKARSTLDRGRNADGDAHSRSPK 127
QY 55 EQREWLSYGCVGVTWNS--GQYPTNRLAFASD-----EDRKNEKNGRPRSGE 104
Db 128 NKREKSTGTASASSQSLEGGYETN-LEIEAIEAPFVODADKEAKLRKKKEEKE 186
QY 105 RAIEFGRVAKESFDEEKGFQREAVSMNRALNAHDESAVLNKKELANGNDALRNE 164
Db 187 RKEREKRELEKKAKAALAAAGAAV-NGNDE--LNSNKSCKSSG-----RRV 239
QY 165 DARSPPYSALRNTPSKERNNGNHD--SRMKAVIYSKHFWSGDRSSADKRYKGGDP 222
Db 240 DPNDPRFAGARTTVTHTMTLTGEIDPVTGRKS-----EYGDID- 278
QY 223 FRPAPGTGLVDMRDRNIPRSPSPGEG-----FVNFYDYGWFG--AOTADADKTVTHG 275
Db 279 ----PNTGDDIPA-----TAVTDPVTGKLILNVAIDPSHFGKQAOVTTTVPITRQ 328
QY 276 NHYHAPNGSLGAMHVYESKFRNWESEYDFDRGA 309
Db 329 QFFD-----GVKHISKGALRRDSEGSDDDMTA 356

RESULT 4
Q9V8R8 PRELIMINARY; PRT; 889 AA.
ID Q9V8R8;
AC Q9V8R8;
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE CORA PROTEIN.
GN CORA OR CG11949.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;

```

Qy	223	FRPAPGTLVDMSDRNIPRSPSPGSG-----FVNFDYGWFG--AQTEADADKTVVTHG	27
Db	556	-----PNTGDIDPA-----TAVTDVPVYTKILINVAQDPSHFQKQAQVQTTEITVPIRQ	605
Qy	276	NHYHAPNGSLGAMHVYSEKFRNWSSEGYSDFDRGA	309
Db	606	QFFD-----GVKHSKGAALRRDSEGSDDDMTA	633
RESULT 5			
ID	Q17240	PRELIMINARY; PRT; 1217 AA.	
AC	Q17240;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)		
DE	SERICIN1B.		
OS	Bombyx mori (Silk moth).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;		
OC	Bombycoidea; Bombycidae; Bombyx.		
FN	NCBI_TaxID=7091;		
OX	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-EUROPEAN 200X300; TISSUE-MIDDLE SILK GLAND;		
RX	MEDLINE-97362906; PubMed-9213370;		
RA	Garel A.A., Deleage G.G., Prudhomme J.J.;		
RI	"Structure and organization of the Bombyx mori Sericin 1 gene and of		
RI	the Sericins deduced from the sequence of the Ser1B cDNA.";		
RL	Insect Biochem. Mol. Biol. 27:469-477(1997).		
DR	EMBL; Z49802; CAA88741.1; .		
SQ	SEQUENCE 1217 AA; 123411 MW; 4C5789F87F6866D0 CRC64;		
Query Match 6.3%; Score 114.5; DB 5; Length 1217;			
Best Local Similarity 20.4%; Pred No. 4.4; Indels 51; Gaps			
Matches 60; Conservative			
Qy	33	NYIRKQOVYSHRDRKQKQMTTEQR---EMLSVCGVTVVWNSGQ-----YPTNR 79	
Db	124	NVSDGQAVASSDARDENRQAQNAQNAQNWADSGYGVSDRSGASSRRRQANYYSKD 183	
Qy	80	LAFASFEDRFKELKNGRPSGSTRAEFEGRVAKESFDEKGFQRAREVASVYMNRALEN 139	
Db	184	ITAKKDSRADSSERSN-----AYNRDSD-----GSENGLSDRSASS 228	
Qy	140	AHDESAYLDNLKKELANGNDALRNEDARSPYSALRNTPSFKERNGNHDPD-----RMK 194	
Db	224	SKNDNVFVYRTKDTI--GGQAKSRSSHSQESDAYYNSPDGYSYNAGTRDSSTSNKKAS 281	
Qy	195	AVIYS--KHFWSGODRSSADKRYGDPDPAPRPAAGTGLVDMSDRNIIPRSPSPGEGFV 255	
Db	282	STIVADKQIIRANDRSSKOLKQSSAQISSGP-KGTSV--SKDKROYNDRKRSKSDAYV 338	
Qy	253	NFDYGWFG--AQTEADADKTVVTHGNHYHAPNGSLGAMHVYSEKFRNWSSEGYS 304	
Db	339	GRD-----GTVAYSNKKDEKTSRQNTNVAQDQNSVRSDDSAASDQTSKSYDRGYS 388	
RESULT 6			
ID	Q9W2X8	PRELIMINARY; PRT; 1161 AA.	
AC	Q9W2X8;		
DT	01-MAY-2000 (TrEMBLrel. 13, Created)		
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)		
DT	01-OCT-2001 (TrEMBLrel. 18, Last annotation update)		
DE	CG15311 PROTEIN.		
GN	CG15311.		
OS	Drosophila melanogaster (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephydroidea; Drosophilidae; Drosophila.		
OX	NCBI_TaxID=7227;		

```

RN 1
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfankoch C., Baidwin D.,
RA Baillet L.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke K., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Flosser C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Correll J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AEO03449; AAF46559.1;
DR Flybase; FBgn0030182; CG15311.
DR InterPro; IPR002106; AA_TRNA_ligase-II.
DR PROSITE; P500339; AA_TRNA_LIGASE-II-2; UNKNOWN.1.
SQ SEQUENCE 1161 AA; 130214 MW; DD01D5B408051D36 CRC64;

Query Match 6.2%; Score 112.5; DB 5; Length 1161;
Best Local Similarity 19.1%; Pred. No. 5.8;
Matches 67; Conservative 66; Mismatches 152; Indels 65; Gaps 13;

QY 8 PPAEPLDRMPDPYRPSYGRAETVNNYIRKQOVYSHRDGRKQOQTEQREWLSCYGV 67
DB 180 PQLEDREQVEHEENAFGR-QSFAYKKLANMHEQQSQQDKRGDDGDE----- 226

QY 68 TWVNSQYTNRLAFASFDEDFR-FKNELKNG-RP--RSGETRAEFEGRVAKESFDEKGFQRA 119
DB 227 VDLDEGEYPPSQL-----FPAIEILNERKFNKPSAGSTGKSRPQLQKKDKQDKMQEQVQ 282

QY 120 EKGFQRAREVASVMNRALNAHDESAYLDNLKKELANGNDALRNEDARSPFYSALRNTPS 179
DB 283 EQQCHKSRQVDEAIGDIENLDNLDELPTNGEDEDGDDADIDDD--EDIKSAIDNDEL 340

QY 180 FKERNGNHDPDMKAVIYSHKFWSGQDRSSADKKRYGDPDAFRPAGTGLVDMSDRN 239
DB 341 AKKYPVATSTTKVPTTLATKTSRSSSSSTTTTMTATSTATSPSPSTTTTK----- 394

QY 240 IPRSPGEG-----FVNFDYGFAGTADADKTVW---THGNH-YHAPNGSLGAM 282
DB 395 -PTPT-PTIGRKLKNFTGLSQPTRYGSDIGRIRAQTQED-DTI-----SEKND 444
```

```

QY 283 GSLGAMHVYESKFRNWSEGYSDFD-----RGAYVITFIPKSNWTAPDKVKQ 328
DB 445 ENVGELHYDISEGSSSRKLVSPDKSEENYLSTYPPGKMN-ATEKKQK 493

RESULT 7
Q65311 PRELIMINARY; PRT; 736 AA.
ID Q65311;
AC Q65311;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE CAPSID PROTEIN.
OS adeno-associated virus 3.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OC NCBI_TaxID=46350;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3H;
RX MEDLINE=96266430; PubMed=8661429;
RA Muramatsu S., Mizukami H., Young N.S., Brown K.E.;
RA "Nucleotide sequencing and generation of an infectious clone of adeno-
RT associated virus 3";
RL Virology 221:208-217(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=3H;
RA Muramatsu S., Brown K.E.;
RA Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RL EMBL; U48704; AAC55049.1;
DR InterPro; IPR001403; Parvo.coat.
DR Pfam; PF00740; Parvo.coat.1.
SQ SEQUENCE 736 AA; 81660 MW; AFF1EF47B5C67A10 CRC64;

Query Match 6.1%; Score 110; DB 12; Length 736;
Best Local Similarity 22.4%; Pred. No. 4.9;
Matches 77; Conservative 49; Mismatches 144; Indels 74; Gaps 22;

QY 18 DYPSPSYGRAETVNNYIRKQOVYSHRDGRKQOQTEQREWLSCYGV--GVTVVNSGQ- 74
DB 4 DGYLPDW--LEDNLSEGIREW---WALKPGVPQKPAKQHQDNRGLVPGYKYLGPNG 58

QY 75 ----YPTNRLAFASFDEDFR-FKNELKNG-RP--RSGETRAEFEGRVAKESFDEKGFQRA 126
DB 59 LDKGEPVNEADAAAEHDKAYDQOLKAGDNPYLKYNHADADEFORLQE---DTSFGNGLG 115

QY 127 REVASVMNRAL--NAHDESAYLDNLKKELANGNDALRNEDARSPFYSALRNTPSFKERN 184
DB 116 RAVFOAKRILLEPLGLVEEAAKTAPGK-----GAYDQSPQEPDSS 156

QY 185 GG----NHDPSPMKAVIYSHKFWSGQDRSSADKKRYGDPDAFRPAP-GTGLVDMSDRN 239
DB 157 SGVGKSGKQPAKREL-----NFGQTGDSVPDPQPLGEPPA---APTSLGNTMASGG 208

QY 240 IPRSPG-----EGFVNFYGFAGTADADKTVW---THGNH-YHAPNGSLGAM 288
DB 209 APMAADNEGAGVNSNGNHCDSQWLGDRIITSTT-WALPTYNHLYKQISSQSGAS 267

QY 289 HVYESKFRNWSE--GYSDFRGAYVITFIPKSNWTAPDKVKQK 330
DB 268 N--DNHYFGYSTPWGYEDFNR--FHCHFSRDMQRL---INNNW 304

RESULT 8
Q07351 PRELIMINARY; PRT; 919 AA.
ID Q07351;
AC Q07351;
DT 01-JUL-1997 (TEMBLrel. 04, Created)
DT 01-JUL-1997 (TEMBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE PROTEASE PRECURSOR.
GN PRTV.
```

OS Vibrio cholerae.  
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
OX NCBI\_TaxID=666;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=017;  
RX MEDLINE=89013889; PubMed=3050359;  
RA Alm R.A., Stroeder U.H., Manning P.A.;  
RT "Extracellular proteins of Vibrio cholerae: Nucleotide sequence of the  
RT structural gene (hlyA) for the haemolysin of the haemolytic EL Tor  
RT strain 017 and characterization of the hlyA mutation in the non-  
RT haemolytic classical strain 569B";  
RL Mol. Microbiol. 2:481-488(1988).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=017;  
RA Manning P.A.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL: Y00557; CAA6836.1; -  
DR MEROPS; M06.002; -  
DR InterPro; IPR000501; PKD\_domain.  
DR Pfam; PF00801; PKD; 2.  
DR SMART; SM00089; PKD; 2.  
DR PROSITE; PS00093; PKD; 1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
KW SIGNAL.  
FT SIGNAL.  
SQ SEQUENCE 919 AA; 101996 MW; 6A80774801FBB9D CRC64;  
  
Query Match 6.0%; Score 109; DB 2; Length 919;  
Best Local Similarity 21.8%; Pred. No. 7.8;  
Matches 67; Conservative 52; Mismatches 111; Indels 78; Gaps 16;  
  
QY 42 YSHR-DGRKQMTSEQREWLSCYGVTVNSGGQYPTNRL-AFASFDED-REFKNEKNGR 98  
DB 446 YSNRGDLKNSRP-----LTIPAGSQTALRFKAWFOIEKDYIARVLNGK 493  
  
QY 99 PRSGETRAEF-----GRVAKESFDEKGFQAR-EVASVMNRALENADH-----ES 144  
DB 494 PIAGNTTWDPFKSLGVLPAISSQSGGWVDAQFDLSAWAGQTVELADYLDGGLAMEG 553  
  
QY 145 AYLDNLKKELANGNDALRNEDARSFYSLRNPFSK--ERNGNHDSRMKAVIYSHF 202  
DB 554 LYVDDLRLEVDGNQTLIDNAEGTSSF-----AFQGTFRKNGGFHE-----ANHY 596  
  
QY 203 WSGQDRSSS-----ADKKRYGDPDAFPAGTGLVDSRDRNIPRSPTSPGEGFVNFY 256  
DB 597 YLLOWSHNDVDOGLANLKRFGQLMSFEPGLLVWYVDESADNWTGK--HPGE----- 647  
  
QY 257 GWFGAGTEADADKTVTHGNHYPHAPNGSLGAMHYVESKFRNWSGYSDFDYGAVYITIP 316  
DB 648 GWLGG-EVDADQNALVMSKTG-----EVAQTRFQVRDATFSLFDQAP--LKLVT 692  
  
QY 317 KSNWNTAPD 324  
DB 693 AGDNTLED 700  
  
RESULT 9  
Q96C72 ID Q96C72 PRELIMINARY; PRT; 401 AA.  
AC Q96C72;  
DT 01-DEC-2001 (TRENBLrel. 19, Created)  
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
DE HYPOTHETICAL 48.6 KDA PROTEIN (FRAGMENT).  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
  
Query Match 6.0%; Score 108; DB 4; Length 558;  
Best Local Similarity 23.8%; Pred. No. 4.7;  
Matches 59; Conservative 31; Mismatches 86; Indels 72; Gaps 11;  
  
RP SEQUENCE FROM N.A.  
RC TISSUE=MUSCLE, AND RHABDOMYOSARCOMA;  
RA Strausberg R.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: BC014625; AAH14625.1; -  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 401 AA; 48615 MW; D9ADFDC029A0851D CRC64;  
  
Query Match 6.0%; Score 108; DB 4; Length 401;  
Best Local Similarity 21.5%; Pred. No. 3.1;  
Matches 68; Conservative 33; Mismatches 118; Indels 98; Gaps 12;  
  
QY 39 QQVYSHRDGRKQMTSEQREWLSCYGVTVNSGGQYPTNRLAFASFDEDREFKNEKNGR 98  
DB 7 EERNLEERKRQKEERR-----ITYREKEEERAEQMLKERERERAERA 57  
  
QY 99 PRSGETRAEFGRVAKESFDEKGFQAR-EVASVMNRALENADHESAYLDNLKELANGN 158  
DB 58 KREELR-EYQERVVKLEEVKRRQRELEIEERRR-----FEERLGD 102  
  
QY 159 DALRNEDAR-----SPFYSALRNTPSFKE-RNGGNHDSRMKAVIYSHFWS 204  
DB 103 SLSRKDSRWGDRDSEGTWRKGPEDSEWRGPPPEKEWRGEGRD----- 147  
  
QY 205 GODRSSADK--RKYGDPD-----AFRP-----APGTGLVDSRDRNIPRSPTSPGEGFVN 253  
DB 148 -EDSRHRDERPRRLGDDDEDFSLRPDDDRVPRGMDDRGPRRGPEDRSRRGADD 206  
  
QY 254 FDYGNFQAOTE-----ADADKTVTHGNHYPHAPNGSLGAMHYVESKFRNWSGYSDFDR 307  
DB 207 DRPSWRNTDDRRPRLIADRGNWRHADDDRRPPRGL-----DEDR 248  
  
QY 308 GAVYITIPKSNWNTAPD 324  
DB 249 G-----SWRTADE 256  
  
RESULT 10  
Q9H0G5 ID Q9H0G5 PRELIMINARY; PRT; 558 AA.  
AC Q9H0G5;  
DT 01-MAR-2001 (TRENBLrel. 16, Created)  
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 66.4 KDA PROTEIN.  
GN DKFZP434K1421.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RX MEDLINE=21154917; PubMed=11230166;  
RA Wiemann S., Weill B., Wellenreuther R., Gassenhuber J., Glassl S.,  
RA Ausorge W., Boecker M., Bloeker H., Bauersachs S., Blum H.,  
RA Lauber J., Duesterhoeft A., Beyer A., Koehler K., Strack N.,  
RA Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,  
RA Wambutt R., Korn B., Klein M., Poustka A.;  
RT "Towards a Catalog of Human Genes and Proteins: Sequencing and  
RT Analysis of 500 Novel Complete Protein Coding Human cDNAs";  
RL Genome Res. 11:422-435(2001).  
DR EMBL; AL136806; CAB66740.1; -  
KW Hypothetical protein.  
SQ SEQUENCE 558 AA; 66390 MW; 99B7BDBCDFD06F98D CRC64;

```

QY   295 FRNWSE--GYSDFRGAYVITFIPKSNWNTADPKVKQG 330
      :  |  ||  ||  :  |  |  :  |  :  |
DB   272 YFGYSTPWGYDFNR--EFCHFSRDMQRL---INNNW 304
      :  |  ||  ||  :  |  |  :  |  :  |

RESULT 12
O56137 PRELIMINARY; PRT; 736 AA.
AC O56137
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, last annotation update)
DE CAPSID PROTEIN VP1.
OS adeno-associated virus 6.
OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Dependovirus.
OX NCBI_TaxId=68558;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94080418; PubMed=9420229;
RA Rutledge E.A., Halbert C.L., Russell D.W.;
RT "Infectious clones and vectors derived from adeno-associated virus
(RAAV) serotypes other than AAV type 2.";
RL J. Virol. 72:309-319(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Rutledge E.A., Russell D.W.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBDJ databases.
DR EMBU; AF028704; AAB95450.1; -.
DR InterPro; IPR001403; Parvo_coat.
DR Pfam; PF00740; Parvo_coat_1.
SQ SEQUENCE 736 AA; 81411 MW; 311217A089C565P5 CRC64;

Query Match          5.9%; Score 106.5; DB 12; Length 736;
Best Local Similarity 21.5%; Pred. No. 8.9;
Matches 74; Conservative 50; Mismatches 147; Indels 73; Gaps 21;

QY   18 DPYRPSYGAEFTWANNYRKMQCVYSHRDGRKQMQTREQRWLSYGCVGVTWNSSGV-- 75
      :  |  ||  ||  :  |  |  :  |  :  |
DB   4 DGLVPDW--LEDNLSEGIREW---WDLKPAPKPKANQQKDDGRLGPLCYKVLGPENG 58
      :  |  ||  ||  :  |  |  :  |  :  |
QY   76 ----PTNLAFASTDEOR-FKNELKNG-RP--RSGETFAEFGEVAKESFDEEKGFORA 126
      :  |  ||  ||  :  |  |  :  |  :  |
DB   59 LDKGEPVNAADAALAHDKAYDQQLKAGDNPLYLRNHADAEEFQEKLQE---DTSEGGNLG 115
      :  |  ||  ||  :  |  |  :  |  :  |
QY   127 REVASYMNPRL--NAHDSAYLDNLKKELANGDALNEDARSFPYSALRNTTPSKERN 184
      :  |  ||  ||  :  |  |  :  |  :  |

```

```
QY   185 GG-----NHDPSPMKAVIYSKHFWSQQDRSSADAKRKYGDPAFRAPAGTGLVDMSRDRI    240
      :|::||:||||:||||:||:||||:||||:||||:||||:||||:||||:||||:||||:
Db   157 SGIGKTGQQPAAKKRL-----NFQGTDGSVPDPOPLGEPPA-TFA-AVGPTTNASGGGA    209
      ||::||:||||:||||:||:||||:||||:||||:||||:||||:||||:||||:
QY   241 PRSPTSFG-----EGEYNFYGFEGAOTEADAOKTYW---THGNHYH--APNGSLGAM     288
      :||:||||:||||:||||:||:||||:||||:||||:||||:||||:||||:||||:
Db   210 PMADNNEGADVGNASGNWHCDSTWLGDRAVTITRT-WALPTYNNHLVKQISSASTGAS     268
      |::||:||||:||||:||:||||:||||:||||:||||:||||:||||:||||:
QY   289 HYESKEPRNWSE--GYSDFDRCAYVIETFPKSNTAPDKVRKGW    330
      |:|:||||:||||:||||:||:||||:||||:||||:||||:||||:||||:||||:
Db   269 N-DNHYGISTFWGYFENR-FCHCFSPDWQRQL----INNWN    305
      ::||:||||:||||:||||:||:||||:||||:||||:||||:||||:||||:||||:

RESULT 13
Q9WBPs PRELIMINARY; PRT; 736 AA.
ID Q9WPB8
AC Q9WPB8
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CAPSID PROTEIN.
OS adeno-associated virus 1.
OC viruses; ssDNA viruses; Parvoviridae; Parvoviriniae; Dependovirus.
```



```
DR EMBL; AF060246; AAD04329.1; -
DR EMBL; AF067399; AAD04342.1; -
DR EMBL; AF060245; AAD04328.1; -
DR EMBL; AF067398; AAD04341.1; -
DR EMBL; AF067397; AAD04340.1; -
DR MGD; MGI:104549; Zfp106.
DR InterPro: IPR001680; WD40.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam; PF00400; WD40; 6.
DR Pfam; PF00096; Zf-C2H2; 1.
DR PRINTS; PR00320; GPROTEINBRPT.
DR SMART; SM00320; WD40; 6.
DR SMART; SM00355; Znf-C2H2; 2.
DR PROSITE; PS50082; WD_REPEATS_2; 2.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 2.
KW DNA-binding; Metal-binding; Repeat; WD repeat; Zinc-finger.
FT CHAIN 556 664 H3A MINOR HISTOCOMPATIBILITY ANTIGEN.
SQ SEQUENCE 1888 AA; 208867 MW; 0216B92854698C58 CRC64;

Query Match          5.8%; Score 105.5; DB 11; Length 1888;
Best Local Similarity 19.6%; Pred. No. 37;
Matches 68; Conservative 37; Mismatches 101; Indels 141; Gaps 17;

QY      8 PPAELDRMPDYPYPSYGRAETVNNYIRKWOQVYSHRDGRKQOQMTBQREWLSYGCVGV 67
DB      109 PPSNSQEVNSDDRQPCWRREDRIPIQDRESYSQPPRRHRG-----PPQDWM----- 154

QY      68 TWVNSQYPTNRLAFASEDEDFKNEKLN-GRPR-----SGETRA----- 106
DB      155 KWEKDG-----FNSTRKNSFPHSLRNSGGPGROSSVYHKGATRGSGSTWFLNHSNSGG 206

QY      107 -----EFGRVAKESFDEE--KGQFQAREVASYMNRALENAHDESAYLNLKKEL 154
DB      207 WHSNNGMVDWNTNGTRNSWHSEGTGGF-----PSWEMNN----- 242

QY      155 ANGNDALRNEDARSPFYSAIRNT-----PSFKERNGNHDPSSRM-----K 194
DB      243 SNGN-----WKS SVRSTNSWNYNGPGDKPQQGRNRNPNYQMEDMTKMANKSNK 291

QY      195 AVIYSKH--FWSGQDRSSSADKKYGDPAFPAPGTGLVDMSRDRNIPRSPTSPGEGFV 252
DB      292 PSKYSQERCKWQRDRDKAAKYR--SPPEGY-----ASDTPFSEGLL 331

QY      253 NFDYGFAGQTEADADKTVWTHGNHYHAPNGSLGAMHYVESKFRNWS 299
DB      332 EFNF----EQRESQTTKQTDTAASKINGKNGTKA-----RDKFRWT 369
```

Search completed: September 27, 2002, 12:40:32  
Job time: 140 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 27, 2002, 12:38:12 ; Search time 13.4 Seconds  
(without alignments)  
956.431 Million cell updates/sec

Title: US-09-884-948-1  
Perfect score: 1811  
Sequence: 1 DSDDRVTPPAEPLDRMPDPY.....ITPIKSWNTAPDKVKQGP 331

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1811	100.0	331	1 TGL_STRSS	P81453 streptococ
2	108	6.0	1382	1 IF3A_HUMAN	Q14152 homo sapien
3	104	5.7	575	1 AMVB_BACCI	P06547 bacillus ci
4	100.5	5.5	926	1 MAY3_SCHCO	P37934 schizosacch
5	99.5	5.5	939	1 ARS2_DROME	Q99587 drosophila
6	99	5.5	2424	1 CCAA_RABIT	P27884 oryctolagus
7	98.5	5.4	1453	1 NKCR_MOUSE	P30415 mus musculu
8	98	5.4	350	1 PHOE_SALTY	P30705 salmonella
9	98	5.4	547	1 BX42_DROME	P39736 drosophila
10	98	5.4	1462	1 NKCR_HUMAN	P30414 homo sapien
11	96	5.3	350	1 PHOE_SALTY	Q56119 salmonella
12	96	5.3	536	1 SKIP_HUMAN	Q13573 homo sapien
13	96	5.3	1344	1 IF3A_MOUSE	P23116 mus musculu
14	96	5.3	2468	1 MAPB_HUMAN	P46821 homo sapien
15	95	5.2	572	1 LMD1_HUMAN	P29536 homo sapien
16	95	5.2	1253	1 MYX9_DROME	Q01989 drosophila
17	94	5.2	462	1 YFHD_HAETN	P44587 haemophilus
18	93.5	5.2	396	1 TITL_DROME	P19351 drosophila
19	93.5	5.2	1722	1 RB22_HUMAN	P29375 homo sapien
20	93	5.1	482	1 UR22_HUMAN	Q15696 homo sapien
21	92.5	5.1	633	1 MLH_TETTH	P40631 tetrahymena
22	92.5	5.1	790	1 KIF9_MOUSE	Q9wv04 mus musculu
23	92	5.1	1101	1 DIA2_HUMAN	Q60879 homo sapien
24	92	5.1	2459	1 MAPB_RAT	P15205 rattus norv
25	91.5	5.1	381	1 NCAP_CVCAI	P36298 canine ente
26	91.5	5.1	383	1 OMS2_SALTY	Q56111 salmonella
27	91	5.0	350	1 R340_ARATH	P2965 arabidopsis
28	91	5.0	535	1 YGHI_CAEEL	Q22836 caenorhabdi
29	91	5.0	636	1 GYRB_THEMEA	P77993 thermotoga
30	91	5.0	1536	1 SIN3_YEAST	P22579 saccharomyc
31	90.5	5.0	351	1 PHOE_KLEPN	P30704 klebsiella
32	90.5	5.0	367	1 OMPC_ECOLI	P06996 escherichia
33	90.5	5.0	980	1 BOB1_YEAST	P36041 saccharomyc

34 90.5 5.0 1131 1 YMS2\_CAEEL P34498 caenorhabdi  
35 90 5.0 351 1 PHOE\_CITFR Q01505 citrobacter  
36 90 5.0 611 1 IF4B\_HUMAN P23588 homo sapien  
37 89.5 4.9 537 1 ARP\_PLAFA P04931 plasmodium  
38 89.5 4.9 935 1 KINE\_SINRA O43093 syncephalids  
39 89.5 4.9 3421 1 TEGU\_HSVB P28955 equine herp  
40 89 4.9 977 1 DLPI\_HUMAN O14490 homo sapien  
41 88.5 4.9 532 1 CCBI\_DROAC O44220 drosophila  
42 88 4.9 544 1 DSKI\_SCHPO P36616 schizosacch  
43 88 4.9 770 1 DAB2\_HUMAN P98082 homo sapien  
44 87.5 4.8 264 1 RS31\_ARATH P92964 arabidopsis  
45 87.5 4.8 303 1 SFRI\_ARATH O22315 arabidopsis

#### ALIGNMENTS

##### RESULT 1

TGL\_STRSS  
ID TGL\_STRSS STANDARD; PRT; 331 AA.  
AC P81453;  
DT 15-DEC-1998 (Rel. 37, Created)  
DT 15-DEC-1998 (Rel. 37, Last sequence update)  
DT 15-DEC-1998 (Rel. 37, Last annotation update)  
DE Protein-glutamine gamma-glutamyltransferase (EC 2.3.2.13)  
DE (transglutaminase) (TGase)  
DE Streptococcus sp. (strain S-8112).  
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;  
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.  
OX NCBI\_TaxID=86037;  
[1]  
RN RP  
SEQUENCE.  
RX MEDLINE=93280110; PubMed=8099353;  
RA Kanaji T., Ozaki H., Takao T., Kawajiri H., Ide H., Motoki M.,  
RA Shimonishi Y.;  
RT "Primary structure of microbial transglutaminase from  
RT Streptococcus sp. strain S-8112";  
RL J. Biol. Chem. 268:11565-11572(1993).  
RC -!- FUNCTION: CATALYZES THE CROSS-LINKING OF PROTEINS AND THE  
CC CONJUGATION OF POLYAMINES TO PROTEINS.  
CC -!- CATALYTIC ACTIVITY: Protein glutamine + alkylamine = protein N5-  
CC alkylglutamine + NH(3).  
KW Transferase; Acyltransferase.  
FT ACT\_SITE 64 64  
SQ SEQUENCE 331 AA; 37862 MW; 5992363A63B2C1FA CRC64;

Query Match 100.0%; Score 1811; DB 1; Length 331;  
Best Local Similarity 100.0%; Pred. No. 4.5e-129;  
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 DSDDRVTPPAEPLDRMPDPYRPSYGRAETVYNNYIRKQQVYSHRDGRKQOQMTDEEQL 60  
Db 1 DSDDRVTPPAEPLDRMPDPYRPSYGRAETVYNNYIRKQQVYSHRDGRKQOQMTDEEQL 60  
Qy 61 SYGCVGVTVNSGGYPTNRLAFASDFEDRKNELKNGRPSGETRAFEGRVAKESFDEE 120  
Db 61 SYGCVGVTVNSGGYPTNRLAFASDFEDRKNELKNGRPSGETRAFEGRVAKESFDEE 120  
Qy 121 KGFORAREVASVMNRALENAHDESAAYLDNLKELANGNDALRNEDARSPTYSALRNTPSF 180  
Db 121 KGFORAREVASVMNRALENAHDESAAYLDNLKELANGNDALRNEDARSPTYSALRNTPSF 180  
Qy 181 KERNGGNDHDSRMKAVLYSKHFWGQDRSSADKKYGDPAEPAPGCTGLVDMSRDNI 240  
Db 181 KERNGGNDHDSRMKAVLYSKHFWGQDRSSADKKYGDPAEPAPGCTGLVDMSRDNI 240  
Qy 241 PRSPTSPGEGFVNFDYGNWFGAQTADAKTWTGHNHYHAPNGSLGAMHYESKFRNWE 300  
Db 241 PRSPTSPGEGFVNFDYGNWFGAQTADAKTWTGHNHYHAPNGSLGAMHYESKFRNWE 300  
Qy 301 GYSDFDGRGAVYITPIPKSWNTAPDKVKQGP 331  
|||||

```

Db 301 GYSDFRGAYVITFIPKSWNTAPDKVKGWP 331
RESULT 2
IF3A_HUMAN
ID IF3A_HUMAN STANDARD; PRT; 1382 AA.
AC Q14152; O00653;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Eukaryotic translation initiation factor 3 subunit 10 (eIF-3 theta)
DE (eIF3 p167) (eIF3 p180) (eIF3 p185).
GN EIF3S10 OR KIAA0139.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96127530; PubMed=9590280;
RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. IV.
RT The coding sequences of 40 new genes (K1AA0121-K1AA0160) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 2:167-174(1995).
RN [1]
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97294683; PubMed=9150439;
RA Scholler J.K., Kanner S.B.;
RT "The human p167 gene encodes a unique structural protein that contains
RT centrosomin A homology and associates with a multicomponent complex.";
RL DNA Cell Biol. 16:515-531(1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97207269; PubMed=9054404;
RA Johnson K.R., Merrick W.C., Zoll W.L., Zhu Y.;
RT "Identification of cDNA clones for the large subunit of eukaryotic
RT translation initiation factor 3. Comparison of homologues from human,
RT Nicotiana tabacum, Caenorhabditis elegans, and Saccharomyces
RT cerevisiae.";
RL J. Biol. Chem. 272:7106-7113(1997).
CC -|- FUNCTION: BINDS TO THE 40S RIBOSOME AND PROMOTES THE BINDING OF
CC METHIONYL-TRNAI AND MRNA.
CC -|- SUBUNIT: EIF-3 IS COMPOSED OF AT LEAST 10 DIFFERENT SUBUNITS.
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- SIMILARITY: BELONGS TO THE EIF3S10 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D50923; BAA09488.1; .
DR EMBL; U58046; ABA1584.1; .
DR EMBL; U78311; AAB80895.1; .
DR MIM; 602039; .
DR InterPro; IPR000717; PCI.
DR InterPro; IPR002017; Spectrin.
DR SMART; SM00088; PINT; 1.
KW Initiation factor; Protein biosynthesis; Repeat; Phosphorylation.
FT DOMAIN 925 1172
FT [CP]-[PS]-[RW]-R-[GN]-[AM].
FT SEQUENCE 1382 AA; 166568 MW; 485C01528D67EBBA CRC64;
Query Match 6.0%; Score 108; DB 1; Length 1382;
Best Local Similarity 21.5%; Pred. No. 2.2;
Matches 68; Conservative 33; Mismatches 118; Indels 98; Gaps 12;

```

```

QY 39 QOYVSHRDGRKQOMTEQREWLSYGCQVTVWNSGOYPTNRLAFASFDDEDRFKNELKNGR 98
Db 785 EERNRLERKQRKQREERR-----ITYYEKEEESQRAEEOMLKREERERAEERA 835
QY 99 PRSGETRAEFEGRAVESDEDEKGFQRAREVASVMNRALENHADESAYLDNLUKKELAN 158
Db 836 KREEELR-EYQERVKKLEEVERKKQRELEIEERERR-----EEERLGD 880
QY 159 DALRNEDAR-----SPFYSLARNTPSFKE-RNGGNHDPFRMKAVIYSKHEWS 204
Db 881 SLSLRKDSRWGDRDSEGTWRKGEADSEWRGPPPEKWRGREGDR----- 925
QY 205 GQDRSSADK---RKVGDDP-----APRGVLVDMSDRNINRPSPTSPGEGFVN 253
Db 926 -EDRSHRRDDEPRRLGDDDEDEPFLRPDDDRVPRGMDDDRGRGPRGDEDRFSRGADD 984
QY 254 FDYGVFGAQT-----ADADKTVWTHGNHYHAPNGSLGAMHYESKFRNWSGEGYDFDR 307
Db 985 DRPSWRNTDDRPRIADEDEGNWRHADDDRPRLGL-----DEDR 1026
QY 308 GAVVITFIPKSWNTAPD 324
Db 1027 G-----SWRTADE 1034
RESULT 3
AMVB_BACCI STANDARD; PRT; 575 AA.
ID AMVB_BACCI
AC P06547;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Beta-amylase precursor (BC 3.2.1.2) (1,4-alpha-D-glucan
DE maltohydrolase).
OS Bacillus circulans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1397;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 11033;
RX MEDLINE=88260890; PubMed=2455212;
RA Siggens K.W.;
RT "Molecular cloning and characterization of the beta-amylase gene from
RT Bacillus circulans.";
RL Mol. Microbiol. 1:86-91(1987).
CC -|- FUNCTION: CATALYZES THE LIBERATION OF MALTOSE FROM 1,4-ALPHA-D-
CC GLUCANS.
CC -|- CATALYTIC ACTIVITY: Hydrolysis of 1,4-alpha-glucosidic linkages in
CC polysaccharides so as to remove successive maltose units from the
CC non-reducing ends of the chains.
CC -|- SUBUNIT: MONOMER.
CC -|- SIMILARITY: BELONGS TO FAMILY 14 OF GLYCOSYL HYDROLASES
CC (BETA-AMYLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y00523; CAA68578.1; .
DR PIR; S03745; S03745.
DR HSP; P36924; I892.
DR InterPro; IPR001554; Glyco_hydro_14.
DR Pfam; PF01373; Glyco_hydro_14; 1.
DR PRINTS; PR00750; BETAAMYLASE.
DR PROSITE; PS00506; BETA_AMYLASE_1; 1.
DR PROSITE; PS00679; BETA_AMYLASE_2; 1.
KW Hydrolase; Glycosidase; Polysaccharide degradation; Signal.

```

```

FT SIGNAL 1 36
FT CHAIN 37 575
FT ACT_SITE 125 125 BETA-AMYLOSE.
FT ACT_SITE 199 199 BY SIMILARITY.
FT ACT_SITE 199 199 BY SIMILARITY.
SQ SEQUENCE 575 AA; 62899 MW; 724E8C0D66B4A258 CRC64;

Query Match 5.7%; Score 104; DB 1; Length 575;
Best Local Similarity 22.9%; Pred. No. 1.5; Indels 116; Gaps 19;
Matches 76; Conservative 37; Mismatches 103;

QY 24 YGRAETVNNYIRKQOVYSHRGRKQOQTEQREWLSCVGVTVW-----NSG 73
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 80 WGYVESAGNQF-DWSYKYVADTVKQ-----AGLKWVPIIISTRGCGNVG 124
QY 74 Q-----YPTNRLAPASDEDFRKNELKNGRPRSG-----ETRAEFEGVAKESDEKGFQ 124
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 DDCNIPSPWLWKSADQFQKDE-----SGYNNESLSPFWGSGVKQ-YDE----- 171
QY 125 RAREVASVMRALENADHESAYLDNLKELANGNDALRNEDARSPYSALRNTPSPFKERN 184
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 172 ---LYASF-----AQFSAKMDMPKIVLSGPP-----SGELRVPYSYPAA 209
QY 185 GGNHDPSPKAVIYKHFWSQDRSSADRKKYGDPAFRPAGTGLVDMRSRDRNIPRSP 244
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 210 GWSY-PARGKPVYTE---TAKSAFTAMTKYGSGLDKINAAGTNLTSMSQ-----ISP 260
QY 245 TSPGEGF-----VNFYDG-----WF-----GAQTEADADKTVVTH-----GNH 277
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 261 PTDSGFGYTGGINITYGKDFLSYQSVLENHLGVIGAAAHKNFDPVFGVRIKAKISGIIH 320
QY 278 YHAPNGSL--GAMHYVESKERNSEGSDFDR 307
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 321 WOMNPSMPSAEH-----AGGYDIYNR 343

RESULT 4
MAY3_SCHCO
ID MAY3_SCHCO STANDARD; PRT; 926 AA.
AC P37934;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mating-type protein A-alpha Y3.
OS Schizophyllum commune (Bracket fungus).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Stereales; Schizophyllaceae; Schizophyllum.
OX NCBI_TaxID=5334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UVM 9-4;
RX MEDLINE=92357793; PubMed=1353886;
RA Stankis M.M., Specht C.A., Yang H., Giasson L., Ullrich R.C.,
RA Novotny C.P.;
RT "The A alpha mating locus of Schizophyllum commune encodes two
RT dissimilar multiallelic homeodomain proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:7169-7173(1992).
CC -1- FUNCTION: SPECIFICS A-ALPHA-3 MATING-TYPE. MAY REGULATE THE
CC EXPRESSION OF GENES SPECIFIC TO THE HOMOKARYOTIC CELL TYPE.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED CONSTITUTIVELY IN HOMOKARYONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M97180; AAB01370.1; -
DR PIR; B37271; B37271.
DR HSSP; P02833; 9ANT.

```

```

DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00046; homeobox; 1.
DR SMART; SM00389; HOX; 1.
DR PROSITE; PS00027; HOMEBOX_1; 1.
DR PROSITE; PS00071; HOMEBOX_2; 1.
KW Homeobox; DNA-binding; Transcription regulation; Nuclear protein.
FT DNA_BIND 147 206 HOMEBOX.
FT DOMAIN 620 660 ARG/LYS-RICH (BASIC).
FT DOMAIN 677 724 SER-RICH.
SQ SEQUENCE 926 AA; 101833 MW; 186631742D3E2E8 CRC64;

Query Match 5.5%; Score 100.5; DB 1; Length 926;
Best Local Similarity 20.5%; Pred. No. 4.9;
Matches 71; Conservative 36; Mismatches 122; Indels 117; Gaps 15;

QY 9 PAEPDLPDPDPYRPSYGRAETVNNYIRKQOVYSHRGRKQOQTEQREWLSCVGVTV 68
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 601 PAEPVIR-PDDFAPVLAELAKRAKRRARKEKKQAEKARK----- 640
QY 69 WYNSGGYPTNRLAFASFDEDFRKNELKNGRPSGTRAEFGRVAKESFDEKGFQARE 128
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 641 -----EKKRKEAKQAKDKRQKAGLPRR-SPTLDS-----SRASS 678
QY 129 VASVMNRALENADHESAYLDNLKELANGNDALRNEDARSPYSALRNTPSPFKERN 188
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 679 VTSADATSRSKRTSRKPRDSSASSVAS-----ARTP---SLSTSS---RRSGTS 724
QY 189 DPS---RMK---AVIYKHFWSQDRSSAD-----KKRYGDPDA----- 222
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 725 MPATPMNESLPVASDNVLTGDKDVTMPELMAQLFGEDDASGLDEPMQSEGSPDWL 784
QY 223 -FRPAPGTGLVDMRSRDRNIPRSPSGEGFYVDFGVWG--AQTADADKTVVTHGNIH 279
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 785 IFSSCNDGALGDMTADVNMPE-----LGLSDTLQSLFDDMMNTSSMDLS 828
QY 280 A-PNGSLGAMHYVESKERNSEGSDFDRGAYVITFIKPSWNTAPD 324
D 1 : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 829 TQPAASFSSSTSSMDFNW-----LLPQCANTAPD 859

RESULT 5
ARS2_DROME
ID ARS2_DROME STANDARD; PRT; 939 AA.
AC Q9V9K7;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Arsenite-resistance protein 2 homolog.
GN CG7843.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.E.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davesport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

```

Dubin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W., Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodak A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Kalish B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D., Merklov G., Milshina N.V., Mobarly C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J.J., Puri V., Reese M.G., Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong E., Sun E., Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
The genome sequence of *Drosophila melanogaster*;  
Science 287:2185-2195(2000).  
-!- SIMILARITY: BELONGS TO THE ARS2 FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF003784; AAF57281.1;  
FlyBase; Fgn0033062; CG7843.  
Hypothetical protein.  
SEQUENCE 939 AA; 106804 MW; 2AD3D528619AC08B CRC64;  
-----  
Query Match 5.5%; Score 99.5; DB 1; Length 939;  
Best Local Similarity 18.3%; Pred. No. 5.9;  
Matches 69; Conservative 43; Mismatches 144; Indels 121; Gaps 13;  
-----  
QY 10 AEPLDRMPDPYRSGRAETVWNNYIRKQOVYSHRDGRKQMTBQREWLVSVCVGVTV 69  
DB 311 ADPVSTQKVPVP-----VNSDGENWDDDAENSAFKKELAEADSKD----- 351  
QY 70 VNSGQYPTNLAFASDEDFKLNELKNGRPR-----SGTRAEFFGRVAKESFDEE 120  
DB 352 --SDSKP-----EDQLNKKTKKRNSSDDSSSSSSSSSDEEKLKRYDYE 399  
QY 121 KGFQREAVSNRALENADHESAYLDNLKELANGNDALRNEDARSPFYSALRNTPSF 180  
DB 400 DGLRAEQKTEAEKDRQATKAKGQPSKLDDEGNEPTEPKGLDSKINTYEIDNLTLS 459  
QY 181 KE-----RNGGNHDPGRM-----KAVIYSKHFWSGODR-----SSADKRYGDPDAFR 224  
DB 460 PEISSNPIKNTDNGDSKVEDEGKPSV-----GKDKVYETETIDLDKVDGQPRALH 512  
QY 225 PAPGTGLVDSROENIPRST-SPGEGFVN-----FDYGFQFCAQTE 264  
DB 513 RTSIFL-----RNLAPSITRSEIAVNCNRFSGYLVAITADPLVRRVTRGWIITFMD 566  
QY 265 ADADKVTWTHGN-----HYHAPNGSLGAMHYES----- 293  
DB 567 VNKEICWGLNNQRLDCMGAIVNRDLRRVRPANGITAHKQVRSRDIKCAKIALNLD 626  
QY 294 -KFRWSEGVDFDRGA 309  
DB 627 EKFLWAEQPKDSDNSA 643

RESULT 6  
CCAA\_RABIT STANDARD; PRT; 2424 AA.  
ID CCAA\_RABIT  
AC P27884; P27883;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Voltage-dependent P/Q-type calcium channel alpha-1A subunit (Calcium channel, L type, alpha-1 polypeptide isoform 4) (Brain calcium channel 1) (Bi).  
DE CACNA1A OR CACNL1A4 OR CACH4 OR CACN3.  
DE Oryctolagus cuniculus (Rabbit).  
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
ME MEDLINE=91187110; PubMed=1849233;  
RA Mori Y., Friedrich T., Kim M.-S., Mikami A., Nakai J., Ruth P., Bosse E., Hofmann F., Flockerzi V., Furuichi T., Mikoshiba K., Imoto K., Tanabe T., Numa S.;  
RA "Primary structure and functional expression from complementary DNA of a brain calcium channel.";  
RT Nature 350:398-402(1991).  
RL [2]  
RN BETA-SUBUNIT BINDING DOMAIN, AND MUTAGENESIS.  
RP MEDLINE=94150724; PubMed=7509046;  
RX Pragnell M., de Waard M., Mori Y., Tanabe T., Snutch T.P., Campbell K.P.;  
RA "Calcium channel beta-subunit binds to a conserved motif in the I-II cytoplasmic linker of the alpha 1-subunit.";  
RT Nature 368:67-70(1994)  
RL [1]  
CC FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION, CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1A GIVES RISE TO P AND/OR Q-TYPE CALCIUM CURRENTS. P/Q-TYPE CALCIUM CHANNELS BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND ARE BLOCKED BY THE FUNNEL TOXIN (FTX) AND BY THE OMEGA-AGATOXIN-IVA (OMEGA-AGA-IVA). THEY ARE HOWEVER INSENSITIVE TO DIHYDROPIRIDINES (DHP), AND OMEGA-CONOTOXIN-GVIA (OMEGA-CTX-GVIA).  
CC -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.  
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.  
CC -!- ALTERNATIVE PRODUCTS: IN THE BRAIN, A SHORT ISOFORM BI-1/1A-1 AND A LONG ISOFORM BI-2/1A-2 (SHOWN HERE), ARE PRODUCED BY ALTERNATIVE SPLICING  
CC -!- TISSUE SPECIFICITY: BRAIN-SPECIFIC. PURKINJE CELLS CONTAIN PREDOMINANTLY P-TYPE VSCC, THE Q-TYPE BEING A PROMINENT CALCIUM CURRENT IN CEREBELLAR GRANULE CELLS.  
CC -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.  
CC -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS FAMILY.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----

-----CAA40715.1; -  
CC ENBL; X57477; CAA40872.1; -  
DR ENBL; X57689; CAA40872.1; -  
DR ENBL; X57478; CAA40744.1; -  
DR ENBL; X57688; CAA40871.1; -  
DR InterPro; IPR000637; Cat\_channel.  
DR InterPro; IPR002077; Ca\_channel.  
DR InterPro; IPR000636; Cat\_channel\_TpL.  
DR InterPro; IPR001682; Channel\_pore\_Ca\_Na.  
DR Pfam; PF00520; ion\_trans; 4.  
DR PRINTS; PR00167; CCHANNEL.  
DR SMART; SM00384; AT\_hook; 1.  
DR Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;  
KW Calcium channel; Glycoprotein; Repeat; Multigene family;  
KW Calcium-binding; Phosphorylation; Alternative splicing.  
FT REPEAT 83 363 I.  
FT REPEAT 473 717 II.  
FT REPEAT 1240 1523 III.  
FT REPEAT 1560 1823 IV.  
FT DOMAIN 1 98 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 99 117 S1 OF REPEAT I (POTENTIAL).  
FT DOMAIN 118 135 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 136 155 S2 OF REPEAT I (POTENTIAL).  
FT DOMAIN 156 167 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 168 185 S3 OF REPEAT I (POTENTIAL).  
FT DOMAIN 186 190 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 191 209 S4 OF REPEAT I (POTENTIAL).  
FT DOMAIN 210 228 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 229 248 S5 OF REPEAT I (POTENTIAL).  
FT DOMAIN 249 335 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 336 360 S6 OF REPEAT I (POTENTIAL).  
FT DOMAIN 361 487 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 488 506 S1 OF REPEAT II (POTENTIAL).  
FT DOMAIN 507 521 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 522 541 S2 OF REPEAT II (POTENTIAL).  
FT DOMAIN 542 549 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 550 568 S3 OF REPEAT II (POTENTIAL).  
FT DOMAIN 569 578 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 579 597 S4 OF REPEAT II (POTENTIAL).  
FT DOMAIN 598 616 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 617 636 S5 OF REPEAT II (POTENTIAL).  
FT DOMAIN 637 689 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 690 714 S6 OF REPEAT II (POTENTIAL).  
FT DOMAIN 715 1253 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1254 1272 S1 OF REPEAT III (POTENTIAL).  
FT DOMAIN 1273 1288 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1289 1308 S2 OF REPEAT III (POTENTIAL).  
FT DOMAIN 1309 1320 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1321 1339 S3 OF REPEAT III (POTENTIAL).  
FT DOMAIN 1340 1350 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1351 1369 S4 OF REPEAT III (POTENTIAL).  
FT DOMAIN 1370 1388 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1389 1408 S5 OF REPEAT III (POTENTIAL).  
FT DOMAIN 1409 1495 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1496 1520 S6 OF REPEAT III (POTENTIAL).  
FT DOMAIN 1521 1575 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1576 1604 S1 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 1605 1609 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1610 1629 S2 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 1630 1637 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1638 1656 S3 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 1657 1665 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1666 1684 S4 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 1685 1703 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 1704 1723 S5 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 1724 1795 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1796 1820 S6 OF REPEAT IV (POTENTIAL).  
FT DOMAIN 1821 2424 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 13 18 POLY-GLY.  
FT DOMAIN 727 732 POLY-GLU.  
FT DOMAIN 1004 1010 POLY-GLY.  
FT DOMAIN 1012 1017 POLY-ARG.

DOMAIN 2219 2227 POLY-HIS.  
DOMAIN 2242 2246 POLY-ARG.  
DOMAIN 2288 2297 POLY-ARG.  
DOMAIN 2298 2301 POLY-GLY.  
DOMAIN 2372 2377 POLY-PRO.  
DOMAIN 2411 2416 POLY-GLY.  
SITE 383 400 BINDING TO THE BETA SUBUNIT.  
SITE 318 318 CALCIUM ION SELECTIVITY AND PERMEABILITY  
SITE 568 568 (BY SIMILARITY).  
SITE 1469 1469 CALCIUM ION SELECTIVITY AND PERMEABILITY  
SITE 1765 1765 (BY SIMILARITY).  
MOD\_RES 1831 1831 PHOSPHORYLATION (BY CAPK) (POTENTIAL).  
CA\_BIND 1849 1860 BY SIMILARITY.  
CARBOHYD 283 283 N-LINKED (GLCNAC. . .) (POTENTIAL).  
VARSPPLIC 772 1051 N-LINKED (GLCNAC. . .) (POTENTIAL).  
VARSPPLIC 772 1051 MISSING (IN ISOFORM CBP107).  
VARSPPLIC 772 1120 MISSING (IN ISOFORM CBP103).  
VARSPPLIC 1857 1884 LYDMVAMLRHMPPLGLGKNCBPVAV -> HYKDMYSL  
VARSPPLIC 2230 2273 RVIQSPVLGLGKCKPVRVAC (IN ISOFORM  
VARSPPLIC 2273 CBP101/CBP109).  
VARSPPLIC 2274 2424 RGPVRSVGSVARRRRRBPVAVRPAPALAHARARAP  
VARSPPLIC 2274 2424 ARK -> PAADAKRIQDPDRPHGHRARDQWRSRPS  
VARSPPLIC 2274 2424 EGREHTTHRQ (IN ISOFORM BI-1/IA-1).  
VARSPPLIC 2274 2424 MISSING (IN ISOFORM BI-1/IA-1).  
VARSPPLIC 2274 2424 A -> T (IN ISOFORM CBS).  
VARSPPLIC 2274 2424 S -> N (IN ISOFORM CBS).  
VARSPPLIC 2274 2424 E -> S; REDUCED BETA-SUBUNIT INTERACTION.  
VARSPPLIC 2274 2424 L -> H; REDUCED BETA-SUBUNIT INTERACTION.  
VARSPPLIC 2274 2424 Y -> S; REDUCED BETA-SUBUNIT INTERACTION.  
VARSPPLIC 2274 2424 E -> A; NO EFFECT ON BETA-SUBUNIT  
VARSPPLIC 2274 2424 INTERACTION.  
SEQUENCE 2424 AA; 273228 MW; F7CC4D0AB4B45604 CRC64;  
Query Match 5.5%; Score 99; DB 1; Length 2424;  
Best Local Similarity 26.5%; Pred. No. 21;  
Matches 43; Conservative 18; Mismatches 51; Indels 40; Gaps 7;  
QY 99 PRSGTAEFEGRVAKESFDEEKQFQRAREVASVMNRALENAHDESAYLNLKKELANGN 158  
Db 828 PQENNNNTNKSVAEPTVDQLGQRAEDFLKQARHHRARDPSAH----- 875  
QY 159 DALRNEDARSPYSA-----LNTSFKE-----RNGNHDPSRKAVIYKHFWSGQDR 208  
Db 876 -AAAGLDARRPWAGSQEALSREGYGRGSHQAREGLEPPG-----FWEGE--- 922  
QY 209 SSSADKRYGDDPAFRPAPGTGLVDMRSRDN-IPRSPTSPGE 249  
Db 923 ---AERGRAGDPHR-RHAHQGVGGSGSRSGSPRTGTADGE 960  
RESULT 7  
NCRC\_MOUSE STANDARD; PRT; 1453 AA.  
ID NCRC\_MOUSE AC P30415;  
DT 01-APR-1993 (Rel. 25, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE NK-tumor recognition protein (Natural-killer cells cyclophilin-  
GN NKTR.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93133824; PubMed=8421688;

RA Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,  
 RA Ortaldo J.R.;  
 RT "A cyclophilin-related protein involved in the function of natural  
 RT killer cells."; Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).  
 RN [2]  
 RN REVISIONS TO C-TERMINUS.  
 RC STRAIN=BALB/C; TISSUE=Blood;  
 RA Anderson S.K.;  
 RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: COMPONENT OF A PUTATIVE TUMOR-RECOGNITION COMPLEX.  
 CC INVOLVED IN THE FUNCTION OF NK CELLS.  
 CC -1- SIMILARITY: CONTAINS 1 CYCLOPHILIN-LIKE PPIASE DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; L04289; AAA37500.2; ALT\_INIT.  
 DR HSSP; Q27450; I433.  
 DR MGI; 97346; Nktr.  
 DR InterPro; IPR0021130; CSA\_PPIase.  
 DR Pfam; PF00150; pro-isomerase; 1.  
 DR PRINTS; PR00153; CSA\_PPIASE; 1.  
 DR PROSITE; PS00170; CSA\_PPIASE\_1; 1.  
 DR PROSITE; PS00172; CSA\_PPIASE\_2; 1.  
 DR Cyclosporin; Isomerase; Rotamase; Repeat; Transmembrane.  
 KW DOMAIN 1 176 PPIASE, CYCLOPHILIN-TYPE.  
 FT DOMAIN 222 241 ARG/LYS-RICH (BASIC).  
 .FT DOMAIN 422 459 ARG/LYS-RICH (BASIC).  
 .FT DOMAIN 964 1003 ARG/LYS-RICH (BASIC).  
 FT DOMAIN 198 273 ARG/SER-RICH.  
 FT DOMAIN 468 565 ARG/SER-RICH.  
 FT DOMAIN 658 842 ARG/SER-RICH.  
 FT DOMAIN 1303 1453 ARG-SER TANDEM REPEAT-RICH.  
 FT SEQUENCE 1453 AA; 163439 MW; DF1173FF814B283E CRC64;  
 -SQ

Query Match 5.4%; Score 98.5; DB 1; Length 1453;  
 Best Local Similarity 17.7%; Pred No.12;  
 Matches 88; Conservative 63; Mismatches 152; Indels 195; Gaps 21;

QY 1 DSDDRVTPP--AEPLDRMPDPYRPSYGRA-----ETVYNNYIRKQQ-- 40  
 Db 355 DDDSETPPHKEEMQL-RAYRPPSGEKWSKGDKLSDPCSSRWDRSLSRQSRWSYNG 413  
 QY 41 VYS-----HRDG-----RKQMTTEOREWLSYCGVGTWYN 71  
 Db 414 YSDLTARSDGHKKHKKRKEKFKKKAKKOKHCHRRHQTAKRR-----IVPDLLE 466  
 QY 72 SGOYPTNRL-----AFASFDEDRFXNELNCRPSRGETRABFEGRVAKE 115  
 Db 467 PRRSPTRHMKSCVRRERRSRASSSSHHSSKRWKSDQDDGSASTHSSRDSY--RSKSH 524  
 QY 116 SPFEKGFQFARAVSMNPALENADESA-----YLDNLKELANGNALNEDAR 167  
 Db 525 SRSDRGSSRSRAVSKSSSSSLNRKSRSSRSGPRRTSISPKPKPAQLSENKPKVTEPLR 584  
 QY 168 -----SPFYSALR----- 175  
 Db 585 PSVPQNGNLVQVAAENIPVILSDSPSPRWKPKQKPKPSYRIQEMKAKTTHLLPV 644  
 QY 176 -----NTPSFKERNGNHDPSRMKAVISKVFWSCQDSSADKKRYCDPDA 222  
 Db 645 QSTYSLTNIKATVSSSSYHKR-----EKPSKESDGSAYSKY-----SDRSSGSSGR--SGSKS 694  
 QY 223 FRPAPGTGLVDMRDRNP-----RSPT-----SPGEGFVNFQYGFQATADAD-- 268  
 Db 695 SRSSSSRSRYTRSRSLPSTRSLRSRSPSRSHSPKNSYDGSQHSRSSSYTSVSSDDGRR 754

QY 269 -----KTWVTEGHNHYHAPNGSLGAMHV-----YESKERNWSEGYSDFD-----RGAYV 311  
 Db 755 AMFRSNKKKSVTSKRRHSNSEKTLHSKYVGRKSSRRHKYSESRSLDYTSDSDSHV 814  
 QY 312 ITFIPKSWNTAPDKVKQG 329  
 Db 815 QVY-----SAPEKEKQG 826

RESULT 8  
 PHOE\_SALTY  
 ID PHOE\_SALTY STANDARD; PRT; 350 AA.  
 AC P30705;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Outer membrane pore protein E precursor.  
 GN PHOE OR STM0320.  
 OS Salmonella typhimurium.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Salmonella.  
 CC NCBI\_TaxID=602;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2;  
 RX MEDLINE=93083994; PubMed=1280609;  
 RA Sierlings G., Elders R., van Lith B., Hofstra H., Tommassen J.;  
 RT "Characterization of the Salmonella typhimurium phoE gene and  
 RT development of Salmonella-specific DNA probes."; Gene 122:45-52(1992).  
 RL Gene 122:45-52(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;  
 RX MEDLINE=21534948; PubMed=11677609;  
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,  
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,  
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,  
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,  
 RA Waterston R., Wilson R.K.;  
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium  
 RT LT2."; Nature 413:852-856(2001).  
 RL Nature 413:852-856(2001).  
 CC -1- FUNCTION: THIS IS ONE OF THE PROTEINS INDUCED WHEN CELLS ARE  
 CC GROWN UNDER PHOSPHATE LIMITATION. ITS PROTEIN PORE IS  
 CC PARTICULARLY EFFICIENT IN THE UPTAKE OF INORGANIC PHOSPHATE,  
 CC PHOSPHORYLATED COMPOUNDS, AND SOME OTHER NEGATIVELY CHARGED  
 CC SOLUTES.  
 CC -1- SUBUNIT: HOMOTRIMER.  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.  
 CC -1- SIMILARITY: BELONGS TO THE OMPC/PHOE FAMILY OF PORINS.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X68023; CAA48164.1; -  
 DR EMBL; AE008709; AAL19276.1; -  
 DR PIR; S25525; S25525.  
 DR HSSP; P02932; 1PHO.  
 DR StyGene; SG10291; phoE.  
 DR InterPro; IPR001702; Gram\_neg\_porin.  
 DR Pfam; PF00267; gram\_ve\_porins; 1.  
 DR PRINTS; PR00182; ECOLNEIPORIN.  
 DR PROSITE; PS00576; GRAM\_NEG\_PORIN; 1.  
 KW Outer membrane; Transmembrane; Porin; Signal; Complete proteome.  
 FT SIGNAL 21 350  
 FT CHAIN 1 350 OUTER MEMBRANE PORE PROTEIN E.  
 FT SEQUENCE 350 AA; 38762 MW; 9D1EE3355AF59877 CRC64;  
 SQ

```

Query Match          5.4%; Score 98; DB 1; Length 350;
Best Local Similarity 20.8%; Pred. No. 2.2;
Matches 64; Conservative 35; Mismatches 116; Indels 92; Gaps 13;

QY 24 YGRAETVYNNYIRKQWQYSHRDGRKQMTTEQREWLSYGCVGVTWNS-----GOYPTN 78
DB 34 YGKVKAM-----HYMSYDSKDG-----DSYVRFKFGKGTQINDQLTGVRWEAE 79
QY 79 RLAFASDEFDRFKNELKNRPSGETRAEFEGRVAKA--SFDEKGFQRAREVASVMNRA 136
DB 80 ---FASNAESDSSOK-----TRLAFAGLKLKDIGSPDYGNLCALYDVAEATDME 128
QY 137 LENAHEDEAYLDNLKELANGDARSPFYSALRNTSPSKERNGNHDPSPRMKAV 196
DB 129 PFGGDSRAQTDNFWTKFASGLAYRNTD-----PFGIVDGLDLILQVCKNEDRD----- 179
QY 197 TYSKHFWSGQDRSSADRRKYGDPAFPAGTGLVDMRDNRNIPRPTSPGEGFVNF 256
DB 180 -VKKONGDGFCTSVSYD-----FSGSD-FAVSGAYTILSDRTREQNLQRRT----- 223
QY 257 GWFGAOTADADKTWTWTHGNHYHAPNGSLGAMHYVESKFRNWSGYSDFDRGAYVITFIP 316
DB 224 ---GDKAE-----WATGVKYDA-----NDIITATFYS 248
QY 317 KSWNTAP 323
DB 249 ETRNMT 255

RESULT 9
BX42_DROME STANDARD; PRT; 547 AA.
AC P39736;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Puf specific protein Bx42.
GN Bx42.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R; TISSUE=Embryo;
RX MEDLINE=95048378; PubMed=1421596;
RW Wieland C., Mann S., von Besser H., Saunweber H.;
RT "The Drosophila nuclear protein Bx42, which is found in many puffs on
RL polytene chromosomes, is highly charged.";
RN Chromosoma 101:517-525(1992).
CC -!- FUNCTION: MAY PLAY A ROLE IN CHROMATIN STRUCTURE AND FUNCTION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- DEVELOPMENTAL STAGE: TWO TRANSCRIPTS ARE DETECTED OF SIZES 1.9 AND
CC 2.2 KB. BOTH ARE DETECTED SOON AFTER FERTILIZATION AND SHOW
CC RELATIVELY CONSTANT EXPRESSION DURING THE FIRST 2/3 OF
CC EMBRYOGENESIS. IN 0-3 HR EMBRYOS, THE SMALLER TRANSCRIPT IS
CC PREDOMINANT AND THE LEVELS OF THE TWO TRANSCRIPTS ARE SOMEWHAT
CC REDUCED AT THE LATER STAGES OF DEVELOPMENT, BUT THEY ARE FOUND IN
CC APPROXIMATELY CONSTANT AMOUNTS DURING LARVAL, PUPAL AND ADULT
CC STAGES. THE SMALLER TRANSCRIPT IS SUSPECTED TO BE A MATERNAL
CC TRANSCRIPT.
CC -!- SIMILARITY: BELONGS TO THE SNW FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

---

```

EMBL; X64536; CAA45834.1; -.
DR FlyBase; FBgn0004856; Bx42.
DR InterPro; IPR004015; SKIP_SNW.
DR Pfam; PF02731; SKIP_SNW; 1.
KW Nuclear protein; DNA-binding. SNW.
FT DOMAIN 177 343
FT DOMAIN 383 462 SH2-LIKE DOMAIN.
FT DOMAIN 209 236 PRO-RICH.
FT DOMAIN 125 155 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 300 400 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 547 AA; 61156 MW; 01399EA291C9D557 CRC64;

Query Match          5.4%; Score 98; DB 1; Length 547;
Best Local Similarity 22.6%; Pred. No. 3.9;
Matches 73; Conservative 38; Mismatches 116; Indels 96; Gaps 19;

QY 44 HRDGRKQOMTEQREWLSYGCVGVTWNSQYD---TNRLA-----FASFD 86
DB 234 HSPSRKTV-KEQKEWKIPPCIS-NMKNAGYTIPLDKRLAADGRGLQVQHINEKFAKVA 291
QY 87 E-----DRFKNELKNRPSGETRAEFEGRVAKESDFDEKGFQRAREVASVMNRALENAH 141
DB 292 EALYIADRKARE-----AVEARSOLEKKLAQK--EKEKKEDMLRMA-----QRAR 335
QY 142 DESAYLDNLKELANGDARSPFYSALRNTSPSKERNGNHDPSPRMKAVIYSKH 201
DB 336 EERA-----GLRNPFAAEPGSGGATGS-EVRRN----- 363
QY 202 FWSGQDRSSADRRKYGDPAFPAP-GTGLVDMSDNRNIPRS-----PTSPGEGFVNF 254
DB 364 ----DLRAERQERQDRNLRQAAPEKRSKLOKEREIDISEQIALGLPAKSAGNETL 418
QY 255 DYGWFGAQTEADA---DKTWTWTHGNHYHAP---NGSLGAMHYVESKFRNWSGYSDFDR 307
DB 419 DQRLNTKGMDSGYGDEAY---NYDKEWRDSNTLGA-HIYRPSKQADSDNYGGDLDA 474
QY 308 GAVVITFIP-KSWNTAPDKVKQG 329
DB 475 IVNTRKRVDPDKQFSGASKEAAG 497

RESULT 10
NKCR_HUMAN STANDARD; PRT; 1462 AA.
AC P30414;
DT 01-APR-1993 (Rel. 25, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE NK-tumor recognition protein (Natural-killer cells cyclophilin-
DE related protein) (NK-TR protein).
GN NKTR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood.
RX MEDLINE=93133824; PubMed=8421688;
RA Anderson S.K., Gallinger S., Roder J., Frey J., Young H.A.,
RA Ortaldo J.R.;
RT "A cyclophilin-related protein involved in the function of natural
RT killer cells.";
RN Proc. Natl. Acad. Sci. U.S.A. 90:542-546(1993).
RP REVISIONS.
RA Anderson S.K.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: COMPONENT OF A PUTATIVE TUMOR-RECOGNITION COMPLEX.
CC INVOLVED IN THE FUNCTION OF NK CELLS.
CC -!- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE VIA ITS N-TERMINUS.
```





```

CC      -!- SIMILARITY: BELONGS TO THE EIF3S10 FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: U14172; AAA90910.1; -
CC      EXEL: X84651; CAA52144.1; -
CC      EMBL: X17373; CAA32246.1; -
CC      PIR: S13800; S13800.
CC      MGD: MGI:95301; Eif3.
CC      InterPro: IPR000717; PCI.
CC      SMART: SM00088; PINT: 1.
CC      Initiation factor; protein biosynthesis. Repeat: Phosphorylation.
CC      DOMAIN 924 1133
CC      CONFLICT 613 647
CC      CONFLICT 683 684
CC      CONFLICT 717 717
CC      CONFLICT 766 766
CC      CONFLICT 787 790
CC      CONFLICT 793 793
CC      SEQUENCE 1344 AA; 161949 MW; F4CAE2169F577712 CRC64;
CC
CC      Query Match 5.3%; Score 96; DB 1; Length 1344;
CC      Best Local Similarity 18.9%; Pred. No. 17;
CC      Matches 53; Conservative 39; Mismatches 103; Indels 86; Gaps 11;
CC
CC      QY 3 DDRVTPAEPLDMPDPRYSYGAETVNNYIRK-WQVYSHRGR-...KQQTTEQ 56
CC      Db 1115 DDRV--PRGGDARPGWPR-...FVKPGWREKEKAREESWGPPRESRPS 1161
CC      QY 57 REWLSYGVGVTVWNSQYPTNRLAFASDFEDRFKFKELKNGRPSRGTAEFEGRVAKES 116
CC      Db 1162 REW-...DRDKEKRDNDRENDKDLDRDRERD 1193
CC      QY 117 FDEKGFQRAEVASVNNRALENHDSAYLDNLKELANGNDALNEDARSFFYSALRN 176
CC      Db 1194 GDREDFRFRPDEGGWRGP-...AESSSWRDSRRDRDRDRDRDR-...RD 1243
CC      QY 177 TPSPFKERNGNHDPSPRMKAVISKH-...FWSGDRSSADKRYGDPDAFRAPGTGLV 232
CC      Db 1244 LRDLDRDLDRDRDRGFLRSEREREASSWRRTD-DRKDRTEERDPPRRVP2PPA-... 1298
CC      QY 233 DMSRDRNIPRSPSPGSGFVNFYWGFGAQTEADAKTYW 272
CC      Db 1299 -LSRDRER-...BREGKEKASW 1318
CC
CC      RESULT 14
CC      MAPB_HUMAN STANDARD; PRT; 2468 AA.
CC      AC P46821;
CC      DT 01-NOV-1995 (Rel. 32, Created)
CC      DT 01-NOV-1995 (Rel. 32, Last sequence update)
CC      DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC      DE Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain
CC      DE LC1].
CC      GN MAP1B.
CC      OS Homo sapiens (Human).
CC      OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC      OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
CC      CX NCBI_TaxID=9606;
CC      RN [1]
CC      RP SEQUENCE FROM N.A.

```

```

RC      TISSUE=Fetal brain;
RC      MEDLINE=95104835; PubMed=7806212;
RC      Lien L.L., Feener C., Fischbach N., Kunkel L.M.;
RC      "Cloning of human microtubule-associated protein 1B and the
RC      identification of a related gene on chromosome 15.";
RC      Genomics 22:273-280(1994).
CC      -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC      THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC      TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC      MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC      STABILIZING MICROTUBULES.
CC      -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC      WITH MAP1A AND MAP1B PROTEINS.
CC      -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC      KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC      responsible for the binding of MAP1B to microtubules.
CC      -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC      from MAP1B by proteolytic processing. It is free to associate with
CC      both MAP1A and MAP1B. It interacts with the amino-terminal region
CC      of MAP1B (BY similarity).
CC      -!- SIMILARITY: TO MAP1A.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
CC      EMBL: L06237; AAA18904.1; -
CC      MIM: 157129; -
CC      InterPro: IPR000102; MAP1B_neuraxin.
CC      Pfam: PF00414; MAP1B_neuraxin; 10.
CC      PROSITE: PS00230; MAP1B_NEURAXIN; 6.
CC      Microtubules; Repeat; Phosphorylation.
CC      CHAIN ? 2468 MAP1 LIGHT CHAIN LC1.
CC      REPEAT 1878 1894 MAP1B 1.
CC      REPEAT 1895 1911 MAP1B 2.
CC      REPEAT 1912 1928 MAP1B 3.
CC      REPEAT 1929 1945 MAP1B 4.
CC      REPEAT 1946 1962 MAP1B 5.
CC      REPEAT 1963 1979 MAP1B 6.
CC      REPEAT 1997 2013 MAP1B 7.
CC      REPEAT 2014 2030 MAP1B 8.
CC      REPEAT 2031 2047 MAP1B 9.
CC      REPEAT 2048 2064 MAP1B 10.
CC      DOMAIN 589 790 LYS-RICH (HIGHLY BASIC, CONTAINS MANY
CC      KKEE AND KKEI/V REPEATS).
CC      SEQUENCE 2468 AA; 270618 MW; 540839C8DF09D461 CRC64;
CC
CC      Query Match 5.3%; Score 96; DB 1; Length 2468;
CC      Best Local Similarity 21.9%; Pred. No. 37;
CC      Matches 68; Conservative 39; Mismatches 110; Indels 94; Gaps 14;
CC
CC      QY 46 DGRKQOMTEEQREWLSYG-...CVGVTVWNSQYPTNRLAFASDFEDRFKFKELKNGRPRS 101
CC      Db 941 DYEKAETEEAEPEDEEGEEHVCVA-...SKHSPTDEESAKADAYIRE-KRESVAS 995
CC      QY 102 GETRAE-...FEGRVAKESFDEKGFQARE-...VASVMNRALE- 138
CC      Db 996 GDDRAEEDMDAELEKGEAEQSEEADEEDKAEAREEVEPEKMEAEYVNAVYDKAAEA 1055
CC      QY 139 -NAHDESAYLDNLKELANGNDALNEDARSFFYSALRNTPSFKERNGNHDPSPMKAVI 197
CC      Db 1056 GAEGSQYGLTPTTKQLG-...AQSP-...GREPASSIHDTLPGG- 1093
CC      QY 198 YSKHFWSGODRSSADKRYGDPDAFRPAPG-...TGLVDMRSDRNIIPRSP 245
CC      Db 1094 -...SESEATASDENREDQPEEFTATSGYTSQSTIISSEPTMDENSTPRDWSDET 1147

```

[illegible]